Inferring Room Geometries

Imperial College
London

Jason Filos

Communications and Signal Processing Group
Electrical and Electronic Engineering

Imperial College London

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

2012 August

Abstract

Determining the geometry of an acoustic enclosure using microphone arrays has become an active area of research. Knowledge gained about the acoustic environment, such as the location of reflectors, can be advantageous for applications such as sound source localization, dereverberation and adaptive echo cancellation by assisting in tracking environment changes and helping the initialization of such algorithms.

A methodology to blindly infer the geometry of an acoustic enclosure by estimating the location of reflective surfaces based on acoustic measurements using an arbitrary array geometry is developed and analyzed. The starting point of this work considers a geometric constraint, valid both in two and three-dimensions, that converts time-of-arrival and time-difference-of-arrival information into elliptical constraints about the location of reflectors. Multiple constraints are combined to yield the line or plane parameters of the reflectors by minimizing a specific cost function in the least-squares sense. An iterative constrained least-squares estimator, along with a closed-form estimator, that performs optimally in a noise-free scenario, solve the associated common tangent estimation problem that arises from the geometric constraint. Additionally, a Hough transform based data fusion and estimation technique, that considers acquisitions from multiple source positions, refines the reflector localization even in adverse conditions.

An extension to the geometric inference framework, that includes the estimation of the actual speed of sound to improve the accuracy under temperature variations, is presented that also reduces the required prior information needed such that only relative microphone positions in the array are required for the localization of acoustic reflectors. Simulated and real-world experiments demonstrate the feasibility of the proposed method.

To my parents

Acknowledgements

First and foremost I would like to thank all my family and in particular my parents for their unconditional mental, emotional and also financial support. Further, I would like to extend my gratitude to my aunt Dr Giannoula Patsantara for always trying to maximise my well-being and my uncle Dr Nikolaos Patsantaras for teaching me what freedom really means.

I would like to thank my supervisor Dr Patrick Naylor who inspired and guided the work in this thesis and for all the suggestions and corrections during the preparation of this manuscript.

I would also like to thank my examiners, Dr Filippo Fazi and Dr Paul Mitcheson for their critical feedback and suggestions.

Many thanks to my colleagues and co-workers at Imperial who were directly linked to this work, Dr Emanuel Habets and Dr Mark Thomas and also Dr Yorgos Tzimiropoulos, Dr Athanasios Gkelias, Dr George Tychogiorgos, Dr Nikolay Gaubitch and Manolis Tsakiris.

Further I would like to thank the group at PoliMi, Prof. Augusto Sarti, Dr Fabio Antonacci and Dr Antonio Canclini and the group at Erlangen, Prof. Rudi Rabenstein and Paolo Annibale.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Introduction

THE word acoustic is derived from the Greek word ἀκουστικός (`akoustikos`), meaning "pertaining to hearing," from ἀκουστός (`akoustos`), meaning "heard, audible," in turn derived from ἀκούω (`akouo`), i.e. the verb "I hear."

Acoustics, and more specifically room acoustics in this work, refers to the study of sound propagation in closed spaces. Fundamentally, the topic of acoustics encompasses the production of sound, the propagation of sound, and the reception of sound. Historically, the study of acoustics is an intrinsically interdisciplinary scientific area, that draws together disciplines such as mathematics, physics, engineering, medicine, psychology, physiology, architecture, general aesthetics, and musical theory.

According to the Physics and Astronomy Classification Scheme (PACS) [4], depicted in Table 1.1, there are three main branches that deal with the study of acoustics, viz., physical acoustics, biological acoustics and acoustical engineering. The main emphasis of the work presented here lies within the latter category, i.e. acoustical engineering, and in particular the studies of room acoustics.

The main contribution of this work, is the development of a novel method for localizing acoustic reflectors, such as walls and furniture, in an acoustic environment, e.g. a conference room, based on acoustic measurements using array signal processing techniques.

Table 1.1: PACS classification system: physical acoustics, biological acoustics and acoustical engineering [4].

Physical Acoustics	**Biological Acoustics**	**Acoustical engineering**
Acoustic theory	Bioacoustics	Acoustic measurements
Aeroacoustics	Musical acoustics	Acoustic signal processing
General linear acoustics	Physiological acoustics	Architectural acoustics
Nonlinear acoustics	Psychoacoustics	Environmental acoustics
Structural acoustics	Speech communication	Transduction
Underwater sound		Ultrasonics
		Room acoustics

Motivated by the echolocation that bats and dolphins use to navigate their surroundings, a framework is established that uses three microphones (or correspondingly four in 3-D) arranged at random (but known) positions inside a room to algorithmically estimate the location of reflectors. In other words, the precise shape of the room is obtained. The relationship between the physical surroundings and the echo-propagation patterns of sound are formulated and analysed from a mathematically geometric viewpoint. The sound emanating from a possibly moving sound source is captured with a set of microphones that are placed at known but arbitrary positions. From the recording of each channel timing information is extracted that relates the propagation patterns to the geometric arrangement of the acoustic scene.

Specifically, a geometric constraint is developed, valid both in two and three-dimensions, that converts time-of-arrival (TOA) and time-difference-of-arrival (TDOA) information into elliptical constraints about the location of reflectors. Multiple constraints are combined to yield the parameters of the line or plane of reflectors by minimizing a specific cost function in the least-squares sense.

Three different scenarios are considered: i) the source and receiver signals are synchronized and the source signal is known, ii) the source and receiver signals are unsynchronized and the source signal is known and iii) the source and receiver signals are unsynchronized and the source signal consists of an unknown impulse-like sounds such as finger-snaps or hand-claps.

When impulse-like signals are used (scenario iii), the estimation of TOAs and TDOAs could be affected by errors, due to the not exactly impulsive nature of the probing signal. TOAs or TDOAs could be affected by errors even in scenario i) and ii), due to non-ideal emission and acquisition systems, thus degrading the accuracy of the localization. In this work the impact of such errors is reduced by employing a template matching filtering technique, which partially compensates for the non-impulsive nature of the probing signal. An iterative localization method is proposed that works well on simulated room impulse responses (RIRs). However, since it considers only a single static source, it is often impossible to obtain a complete set of TOAs, i.e. a set that will contain TOAs of all reflectors, in real acoustic environments. An alternate approach, also proposed in this thesis, considers a space parametrization based on the Hough transform and the data fusion of multiple sequential measurements, that is practical for real environments in which higher-order reflections can arrive at the microphones before the first-order reflections (e.g. corridor-style rooms) or when first-order reflections coming from different walls are hardly distinguishable in the impulse response (e.g. square rooms). By considering acquisitions from multiple source positions, the benefit of this method is studied in detail in terms of improved localization accuracy.

The methods outlined in this work are evaluated both using simulations and real-world experiments. In both cases the proposed approach exhibits low errors with up to millimetre accuracy. A theoretical analysis of the propagation errors is also given.

The rest of the thesis is organized as follows: the remainder of this chapter outlines the historical development of the scientific field of acoustics along with an overview of related state-of-the-art approaches in room acoustics, array signal processing and geometry reconstruction. Chapter 2 outlines relevant background material regarding system identification, room acoustics, sound source localization and line estimation techniques. Chapter 3 introduces the problem formulation and the system model along with the geometric constraint for reflector localization. Chapter 4 considers the problem of reflector localization in two-dimensions and introduces the estimator along with a theoretical error propagation analysis and experimental results. In Chapter 5 the reflector localization estimator is described for the three-dimensional case along with the appropriate error propagation analysis and experimental verification. Chapter 6

discuss potential applications and suggest directions for future work. Finally, Chapter 7 summarizes the methods used in this work.

Relevant Publications

1. P. Annibale, F. Antonacci, P. Bestagini, A. Brutti, A. Canclini, L. Cristoforetti, E.A.P. Habets, J. Filos, W. Kellermann, K. Kowalczyk, A. Lombard, E. Mabande, D. Markovic, P.A. Naylor, M. Omologo, R. Rabenstein, A. Sarti, P. Svaizer and M.R.P. Thomas, "The SCENIC Project: Space-Time Audio Processing for Environment-Aware Acoustic Sensing and Rendering," *the Proc. of the Audio Engineering Society (AES) 131st Convention*, New York, USA, October 20 - 23, 2011.

1.1 History of Acoustics

As noted by Robert B. Lindsay in his historical account of the science of acoustics in [5], the insight of "the evolution of the concepts basic to a given branch of science can often suggest useful ways of approaching current experience." It is therefore of vital importance to analyse the roots and evolution of the science of acoustics in order to place the concepts presented in this manuscript into context.

According to [5], acoustics occupy a "somewhat anomalous position in the hierarchy of the sciences." The scientific area of acoustics, perhaps closely resembling the school of musical thought, does not deal with a continuous and clearly marked out body of source material and is "confronted instead with a bewildering diversity of ideas" [6]. In Ancient Greece, the study of music[1], and consequently acoustics, was part of the quadrivium[2] consisting of arithmetic, geometry, astronomy and music. It can therefore be seen as an interdisciplinary science, closely related to the other three mathematical disciplines. The history of musical thought and acoustics coincide in part with numberless branches of historical study, including those more general disciplines such as cultural history, the history of philosophy, and intellectual history [6]. This complex

[1] Music can be understood in this context as `musica scientia`, i.e. the science of music, or in other words acoustics.

[2] The word is Latin, meaning "the four ways" and was referred to as such during the Renaissance Period.

state of affairs is "undoubtedly due to the peculiarities of music itself; existing in intimate relationship with almost every aspect of human activity, music could hardly give rise to an isolated or well-defined area of thought" [6]. Acoustics and music are intrinsically linked with historical and non-historical disciplines, subjective and analytic and systematic ideas, and, perhaps like the design and construction of musical instruments, is not only a "technical or scientific discipline but a sort of craft, or an 'art' in the classical meaning of this word" [7]. During the past century, however, acoustics, and more specifically room acoustics, have become an exact science. Consequently, the work presented in this manuscript follows the practice of rigorous scientific analysis and does not ponder the different historical interpretations of music and acoustics over the centuries. The analogies to 'art' mentioned previously merely serve as a hint into the intricacies surrounding the historical development of music and acoustics.

Historically speaking, the problems of acoustics are most conveniently divided into three main groups: viz., the production of sound, the propagation of sound, and the reception of sound [5]. The following two sections closely follow Lindsay's historical overview but expand on certain areas of the material given in [5] while omitting other aspects enumerated in that work, most notably pertaining to the developments on the reception of sound.

1.1.1 Production of Sound

It is usually assumed that the first Greek philosopher to study the origin of musical sounds was Pythagoras of Samos (Πυθαγόρας ὁ Σάμιος) who established his school in Crotone in southern Italy in the 6th century B.C. According to the knowledge in the antiquity and the writings on music of the Roman philosopher Boethius in the 6th century A.D., it was Pythagoras who discovered the relationship between musical notes and mathematical equations. This radical discovery paved the way for a scientific approach of quantifying music using mathematical principles [8] and as claimed by Werner Heisenberg ranks amongst the strongest impulses of scientific thought [8]. The relationship between pitch and the frequency of vibration of the sound-producing object in a modern scientific context dates back to the work of Galileo Galilei published in 1638 [9]. However, it is often argued that Galileo's contribution to the mechanics of vibration has been exaggerated and that Isaac Beeckman had published his work on the vibration of strings as early as 1618 [5]. Even earlier than Beeckman was the Italian,

Giovanni Battista Benedetti, who in a work on musical intervals, published in Turin in 1585, stated the equality between the ratio of pitches and the ratio of the frequencies of the vibrating motions corresponding to the production of the sounds. More elaborate studies, based on empirical observations of the vibrations of a stretched string, were published by Marin Mersenne in 1625 [10].

Robert Hooke later experimentally showed the connection between the frequency of vibration and pitch [11], further deepened by the studies of Joseph Sauveur in his seminal work published in 1701 [12]. Sauveur also experimented with strings but the first to provide a strictly dynamical solution of the vibrating string model was Brook Taylor in [13]. Fundamental to the progress of acoustics as an exact science at the time was a mathematical description of the motion of continuous media. This breakthrough came in 1822 with J. B. J. Fourier's analytical theory of heat that described the possibility of expressing any arbitrary function in terms of an infinite series of sines and cosines [14]. The problem of the vibrating string was also studied by J. L. Lagrange in the 18th century and his model was subsequently adopted by J. W. Strutt Lord Rayleigh in his seminal work *Theory of Sound* [15, 16] that continues to live on in most modern textbooks on the subject of mechanics and acoustics [5].

The mathematical scientists of that century of course realised that other solid bodies besides strings emitted sound when disturbed. The theory of elasticity, including elastic vibrations giving rise to sound, were studied by Leonhard Euler and Daniel Bernoulli as early as 1734 [17]. The mathematical methods they used were later systemised and extended by Lord Rayleigh in [15, 16]. The corresponding analytical solution for the vibrations of a solid elastic plate were published by E. F. F. Chladni in 1787 [18]. Some exact forms used in the work by Chladni defied analysis for many years and it was not until 1850 that G. R. Kirchhoff gave a more accurate theory [19]. To this day, studying the vibrational pattern of an airplane fuselage, for example, still supports research on the vibrations of plates and solid shells of various shapes.

The ability to excite vibrations in media of arbitrary nature, size, and shape and with arbitrary frequency over a wide range had to await the development of electroacoustics [20], largely a product of 20th-century research [5]. The historical account of the production of sound continues at this point with the evolution of electroacoustics and the contributions of Lord Rayleigh that mark the end of the classical era in acoustics and the beginning of the modern age of sound.

1.1.2 Propagation of Sound

From the earliest scientific records, dating back to the Greek philosopher and polymath Aristotle (Ἀριστοτέλης) of the 4th century B.C., the predominant idea of the propagation of sound relied on some activity of the air from one point in space to the other, i.e. involving actual motion of air. Aristotle understood that sound consisted of contractions and expansions of air but failed to realize that the sound velocity was independent of the pitch of the sound and consequently taught his following that high pitched notes were transmitted faster than low pitched notes [5]. The Roman architectural engineer, Marcus Vitruvius Pollio of the first century B.C., compared the wave theory of sound to analogies that he drew with surface waves on water. Vitruvius wrote a treatise on the acoustic properties of ancient theaters that included a discussion on interference, echoes, and reverberation, and is credited for this as the forefather of modern architectural acoustics.

Even during the Galilean period, Pierre Gassendi, [10], attributed the propagation of sound to the emission of a stream of "very small, invisible particles from the sounding body, which, after moving through the air, are able somehow to affect the ear" [5]. During that time, and until the 18th century, Gassendi [10], Mersenne [10] and W. Derham [21], along with a group of scholars commissioned by the Academy of Sciences of Paris, conducted experiments to estimate the propagation velocity of sound in air as a sound-transmitting medium.

The measurement of the velocity of sound in solid media was first considered by Chladni and the velocity of sound in liquids by Daniel Colladon in 1826. Although the propagation of sound through air had already been compared with the motion of ripples on the surface of water as early as the first century B.C. by Vitruvius, the first attempt to theorize, in a mathematical form, the wave theory of sound came from Sir Isaac Newton in his famous *Principia mathematica* [22]. Newton compared the transmission of sound with pulses produced by a vibrating body moving the adjacent portions of the surrounding medium that in turn move those adjacent to themselves and so forth. The derivations of Newton earned heavy criticism by contemporaries such as Euler, John Bernoulli and Lagrange and it became clear "that the problem of sound propagation would never be completely solved until the wave equation for sound waves in a fluid could be set up and solved" [5]. An analogy to the equation for a continuous string,

but applied to sound waves, was given by Euler, who laid the foundation for the theory of the propagation of sound waves in air. During the same time, Lagrange, in turn, revised Newton's reasoning and generalized it to the case of sound waves of arbitrary character as distinct from simple harmonic waves [5].

The propagation of a compressional wave in a three-dimensional fluid medium was first considered by S. D. Poisson in 1819 [23] but it remained, however, for Hermann von Helmholtz to give in 1860 "a more thorough treatment of this whole problem" [5].

The problem of the reflection and transmission of a plane sound wave incident obliquely on the boundary of two different fluids was solved by George Green in 1838 [24]. Note that all equations for the wave propagation up to this point are assumed to be linear. It was not until the middle of the 19th century that nonlinear acoustic-wave propagation was taken up by the German mathematician Georg F. B. Riemann [25] and British mathematician and physicist S. Earnshaw [26] who both "more or less independently investigated certain special cases" [5].

The historical account of the propagation of sound continues at this point with the development of the theory of shock waves and the contributions of G. G. Stokes, J. Challis, W. J. M. Rankine, H. Hugoniot and Lord Rayleigh, amongst others. The interested reader is referred to [27] for detailed bibliographical references of the aforementioned authors and the 20th-century development on the subject of nonlinear acoustics.

1.2 Literature Review

Following the historical overview on the development of acoustics over the centuries, the aim of this section is to summarise related research on the subject of modern acoustics with particular emphasis placed on research related to room acoustics. Broadly speaking, the cutoff point between what is considered modern era acoustics and the classical era, is marked by Lord Rayleigh's amalgamation of knowledge in the two volumes of the *Theory of Sound* [15, 16] published in 1894 and 1896 respectively. The work described in these two volumes is still relevant to this day, and as pointed out above, marks the end of the classical era of acoustics and the dawn of modern acoustics.

1.2.1 Room Acoustics

The problems imposed on an audio system by the local environment in which it is employed have long been acknowledged [28, 29]. One of the earliest publications related to modern-day room acoustics is Wallace C. Sabine's groundbreaking work on architectural acoustics [30] published in 1922. Sabine gathered extensive amounts of data on existing acoustic spaces and from this information was able to derive empirical guidelines to aid future design work. He is credited for the development of the theories of reverberation leading to an equation for calculating the reverberation time that is ultimately linked to the shape of the room. This was in part aided by the derivation of the equation of decaying sound in a room by W. S. Franklin [31].

By 1930 the method of images was used to perform an analytical study of the decay of sound energy in a room [32]. This approach, where enclosure boundaries are treated as 'acoustic mirrors,' has been used extensively in the search for a "greater understanding of the requirements of a space in which sound is to be controlled" [33]. The first pieces of research discussing the image-source approach for calculating sound fields in simple rectilinear enclosures were given by Gibbs and Jones [34] and Allen and Berkley [35] in 1972 and 1979 respectively. The image-source model was later extended to the case of arbitrary polyhedra [36] and higher-order reflections [37].

An alternative, but closely related, approach called ray tracing, where sound 'rays' are considered to radiate from a source in all directions, was established following the development of the method of images [38]. Examples of publications that estimate the acoustical room response using ray tracing techniques are [39, 40].

More recently, image theory and ray theory methods were extended by finite-element analysis that allow the modeling of more complex geometries, non-uniform absorption and multiple sources [33, 41]. A more detailed overview of these studies is given in Chapter 2.

Having described the simulation, or prediction, of room impulse responses, focus is shifted towards the measurement of such room impulse responses, that are also referred to as room transfer functions in this context. A comparative study of the various measurement techniques is given in [42] that includes methods such as maximum-length sequences and time-stretched pulses [43–48]. Additionally, other related approaches, that include the modeling and signal processing of the sound source and the fusion of

simulated or measured visual cues together with acoustic measurements, are given in for example [41, 49–54].

Ultimately, both the simulation and measurement of the response of a room lead to a better understanding of the effects the environment imposes on the wave propagation of source signals and the behaviour of space-time processing algorithms such as acoustic echo cancellers, speech enhancement algorithms or dereverberation algorithms. It is often impossible to take such knowledge into account in an *a priori* way, such as when the acoustics of an existing room need to be improved after its construction. Evidently, the acoustic behaviour of an enclosure should be taken into account by the architect before construction [55–57], however it is often important to achieve acoustic control of an existing acoustic space by electroacoustic means. Such approaches are important for repairing disturbing deficiencies in the acoustics of a given enclosure.

These approaches are based on "acoustic holography" where the sound field is reconstructed using direct and reflected wave fields with desired wavefront properties at each moment of time [58]. Ultimately, these reconstructed sound fields cannot be distinguished from true sound fields. Examples of such wave field synthesis approaches are given in [58–62] and were first developed in 1988 by A. J. Berkhout. Contrary to traditional spatialization techniques such as stereo or surround sound, the localization of virtual sources in wave field synthesis does not depend on or change with the listener's position. Other room-loudspeaker control techniques are generally referred to as local room equalization [29].

1.2.2 Array Signal Processing

Sensor array signal processing emerged as an active area of research as a consequence of the second world-war. It was centered on the "ability to *fuse* data collected at several sensors in order to carry out a given estimation task (space-time processing)" [63]. The first such space-time processing tasks on an array of sensors were based on spatial filtering or beamforming and later adaptive beamforming and time delay estimation techniques [63]. The extension of the time-delay estimation methods to more than one signal and the limited resolution of beamforming renewed interest in statistical signal processing [64] and the emergence of the parameter estimation approach as an active area of research [63]. Early applications focused on maximum likelihood estimators (MLE) [65, 66]. The introduction of subspace-based estimation techniques, [67] such

as MUSIC [68] and ESPRIT [69–71] that separate the signal and noise subspaces, marked the beginning of a new era in the sensor array signal processing literature [63]. Alternative tracking and localization algorithms emerged, particularly in the area of radar and sonar signal processing, suitable for tracking dynamic signals. These include approaches based on Kalman filtering [72–76], particle filtering [77] and even H_∞ filters [78].

Delay estimation problems related to room acoustics using sensor arrays of arbitrary form generally fall in two categories: viz. likelihood-based and least-squares using either iterative or closed-form estimators [79, 80]. The first least-squares approach that solved the hyperbolic fix problem [81], based on range difference measurements, appeared in 1972 using an estimator called plane intersection [82]. Maximum likelihood methods for time delay estimation grew in popularity because of the proven asymptotic consistency and efficiency of the MLE [83–85]. The derivations of the Cramer-Rao lower bound for some of these delay estimation problems are presented in [85, 86]. Unfortunately, because the number of microphones in an array is usually limited, the MLE is not optimal since the large-sample problem turns into a finite-sample problem [80].

Later, closed-form estimators [87–89] gained wider attention because of their applicability for real-time applications [80]. On basis of the plane intersection estimator a series of highly effective closed-form least-squares estimators appeared such as the spherical-interpolation (SI) method [90–92]. However, these SI estimators were not optimal in the least-squares sense. A noniterative SI estimator that approximated the MLE was proposed in 1994 by Chan and Ho [93]. This method improved the SI estimation with a second LS estimator. A linear-correction LS approach [80] appeared afterwards along with other constrained LS estimators [94]. Sensor placement and acoustic source estimators relevant to room acoustics were studied in [95–97].

1.2.3 Geometry Reconstruction

Inferring the geometry of a room deals with the problem of localizing acoustic reflectors, such as the walls or furniture, in an enclosure using acoustic measurements. Geometrical acoustic methods appeared relatively recently in the research literature. First examples did not make use of acoustic signals in the audible range but focused on ultrasonic transducers [98]. Often, more important than the estimation of the geometry of the acoustic enclosure were the acoustic parameters of the room [99, 100].

Some initial room geometry reconstruction and analysis approaches relied on visual cues such as visual cameras [101] or a mixed approach using both visual and audio cameras [102]. The first room modeling or acoustic scene reconstruction methods using controlled emissions appeared only in this decade [103]. The authors in [103] made use of a loudspeaker rotating on a circular pattern emitting a controlled noise that was in turn captured by a microphone in the centre of the circle. This allowed the inference of reflectors (one at a time) using a likelihood map. Alternative methods using microphone arrays [104, 105] and spherical microphone arrays [106] followed. These were most commonly based on beamforming approaches [107]. The authors in [108] proposed the estimation of reflective surfaces from continuous signals as an alternative to room impulse response measurements but later also adopted the impulse response based approach [109]. Interestingly, the authors in [110] claimed that it was possible to reconstruct the geometry of an acoustic enclosure from a single acoustic impulse response (AIR), which is in contrast to most of the aforementioned works. While this is indeed true for most convex polygonal geometries there are certain modeling assumptions that are not fully satisfiable. The key point in this paper is that the source and sensor are co-located. It turns out to be a straightforward problem in this very special case because one of the unknowns (the relative displacement of source and sensor) is eliminated. The work in [110] was later extended in [111, 112].

A geometric approach on which this manuscript is based on, for localizing acoustic reflectors in a room, was first proposed in [113]. A single microphone and a moving sound source were used to obtain the time of arrivals (TOAs) of the reflections assuming that the source and receiver were synchronized. The TOA information was used to form a set of elliptical constraints on the possible locations of the reflector. The common tangent of these constraints was shown to correspond to the reflector location that can be found by minimizing a specific cost function in the least-squares sense. In [114] the authors considered the case when measurements were unsynchronized. In this case, the TOAs cannot be found directly, and an additional step was proposed that estimates TOAs from the TDOA estimates. A technique for the estimation of multiple reflectors from a single set of measurements was also proposed, that iteratively minimizes a global solution space. More recently, the authors proposed a robust inference method [115] utilizing a closed-form solution to minimize the cost function. In addition, a parametrization based on the Hough transform was introduced that increases

the robustness to errors in the TOA estimation. This work was extended in [3, 116] to derive an exact solution for the minimization of a cost function combining an arbitrary number of quadratic constraints and additionally outlining a method for the analytic prediction of reflector localization accuracy.

The development of the geometric approach, outlined in the previous paragraph, considers the problem of reflector localization in two-dimensions. An extension of the geometric approach to three-dimensions was first proposed in [117]. This approach in turn was complemented in [118] with the Hough transform parametrization and finally in [119] with an exact minimization approach.

1.3 Statement of Originality

The author believes the following points represent original contributions:

- A geometric constraint, formulated using linear algebra, that converts TOA and TDOA information into elliptical constraints about the location of reflectors. The ellipse parametrization contains a mathematical proof not found in the scientific literature to date.
- The removal of necessary synchronization between source and receivers, due to the estimation of the source position as a first step, and the subsequent conversion of TDOA information into TOAs.
- The estimation of multiple reflectors, present in the acoustic environment, from only one acoustic measurement that gives the exact solution using simulated RIRs.
- A robust inference method utilizing a closed-form solution to minimize the cost function along with a parametrization based on the Hough transform that localizes multiple reflectors and increases the robustness to errors in the TOA estimation when using real-world data.
- An extension of the geometric approach to three-dimensions by combining multiple two-dimensional orthogonal line parameters.
- An analytical prediction of the impact of errors on measurements on the reflector localization error both in two and three-dimensions.
- An application of the geometric inference framework to environments that are exposed to strong temperature variations. Variations in the ambient temperature impair the accuracy of the TOA estimation process due to local variations in the speed of sound.

2

Background

Introduction

THE purpose of this chapter is to introduce the concepts of system identification, simulation and measurement of room impulse responses, source localization and line detection. The former three concepts are developed from a signal processing perspective while the latter is based on a computer vision/image processing paradigm.

In the system identification process, the single-input multiple-output (SIMO) case is considered, adhering to a formulation that is related to speech processing in reverberant environments. Broadly speaking, four different classes or semantics of system identification are used throughout this work, viz., i) supervised identification using controlled input signals (e.g. MLS, TSP); ii) supervised identification using uncontrolled input signals (e.g. AEC algorithms); iii) unsupervised identification using controlled input signals (e.g. BSI algorithms such as NMCFLMS); iv) unsupervised identification using uncontrolled input signals (e.g. using an impulsive source signal, such as the snapping of fingers, to directly estimate an impulse response by using a matched filter approach that is not, strictly speaking, system identification but more of a blind deconvolution approach aimed at improving the temporal spread caused by the unknown excitation signal).

As a next step, the mathematical formulation of the wave equation is introduced. The solution to the wave equation yields an impulse response as calculated between a static sound source and a receiver. However, due to the analytical form of the wave

equation, numerical methods need to be used to simulate a Room Impulse Response (RIR) in a computer environment. An overview of different approaches for simulating such RIRs is given and their appropriateness is discussed with respect to the methods established in this work.

The concept of source localization is then introduced, based on two existing approaches found in the scientific literature. The source localization methodology can be used to remove a priori assumptions about the estimated RIRs.

Finally, a line parameter estimation methodology is introduced, based on an approach originally developed in the field of image processing. Estimating the parameters of a line segment by using the coordinate points or pixels of an image is achieved by means of the Hough transform.

The chapter is organized as follows: Section 2.1 outlines the system identification problem, where we distinguish between the supervised and unsupervised identification case. Section 2.2 introduces the wave equation and discusses methods for approximating the analytical solution. Particular emphasis is drawn towards the method of images, for simulating RIRs and methods for obtaining Acoustic Impulse Responses (AIRs) in a real room. Section 2.3 outlines the source localization problem, that can be used to identify time-of-arrivals (TOAs) from unsynchronized measurements by considering the more readily available time-difference-of-arrival (TDOA) information calculated between sensor pairs. Finally, in Section 2.4 the Hough transform is introduced.

2.1 System Identification

System identification is the process of establishing a mathematical model of an unknown dynamic system by analyzing the relationship between its input and output data. This problem is at the core of a large variety of signal processing and communications applications. Identification essentially means developing an understanding of how the input signal is transmitted, processed and distorted by the system [120, 121].

Since in practice the audio channel, representing the wave propagation of sound inside an acoustic enclosure, is generally non-stationary and usually has a long impulse response, determining its characteristics is not easy, even when the input signal is known, such as in the case of acoustic echo cancellation. However, in many other cases, e.g., acoustic dereverberation, wireless communications, time delay estimation, etc.,

the input is either unobservable (i.e. it cannot be seen by the signal processor) or very expensive to acquire; the choice inevitably comes down to a blind method, and as a result, the channels are more difficult to estimate [121].

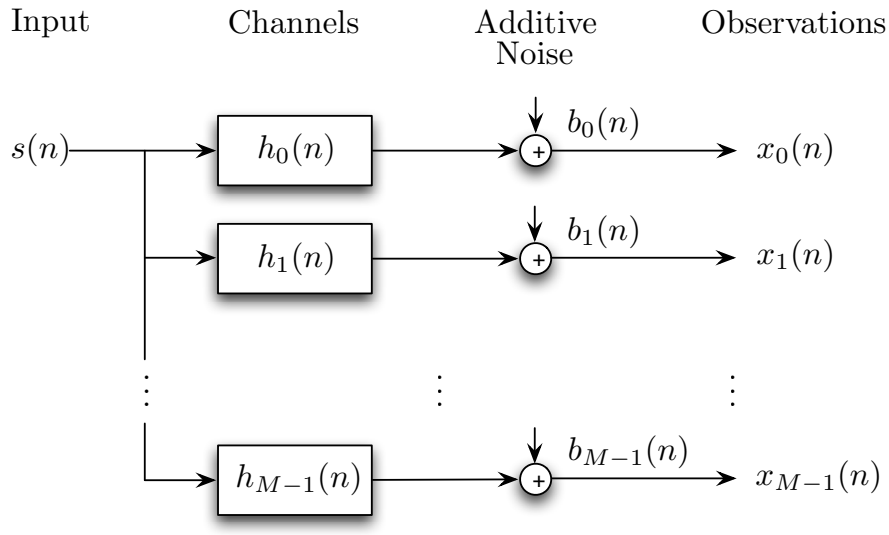


Figure 2.1: Illustration of the relationships between the input $s(n)$ and the observations $x_i(n)$ in a single-input multiple-output FIR system.

Consider a Single-Input Multiple-Output (SIMO) Finite Impulse Response (FIR) linear system as shown in Fig. 2.1. The ith channel output signal $x_i(n)$ at sample n is the result of a linear convolution between the source signal $s(n)$ and the corresponding channel impulse response h_i, corrupted by an additive background noise $b_i(n)$:

$$x_i(n) = h_i * s(n) + b_i(n), \quad i = 0, 1, \cdots, M-1, \tag{2.1}$$

where M is the number of observations (i.e. channels) and the additive noise is assumed to be zero-mean and uncorrelated with the source signal. The ith impulse response with L coefficients can be expressed as

$$\mathbf{h}_i(n) = [h_{i,0}(n)\ h_{i,1}(n) \cdots h_{i,L-1}(n)]^T, \quad i = 0, 1, \cdots, M-1, \tag{2.2}$$

Since the the impulse responses are assumed to be slowly time-varying, $\mathbf{h}_i$ is independent of n [122]. The ith sensor signal (2.1) can be written in vector form as

$$\mathbf{x}_i(n) = \mathbf{H}_i \mathbf{s}(n) + \mathbf{b}_i(n), \tag{2.3}$$

where

$$\begin{aligned}
\mathbf{x}_i(n) &= [x_i(n)\ x_i(n-1)\ \cdots\ x_i(n-L+1)]^T, \\
\mathbf{H}_i &= \begin{bmatrix} h_{i,0} & \cdots & h_{i,L-1} & \cdots & \cdots & 0 \\ 0 & h_{i,0} & \cdots & h_{i,L-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & h_{i,0} & \cdots & h_{i,L-1} \end{bmatrix}, \\
\mathbf{s}(n) &= [s(n)\ s(n-1)\ \cdots\ s(n-2L+2)]^T, \\
\mathbf{b}_i(n) &= [b_i(n)\ b_i(n-1)\ \cdots\ b_i(n-L+1)]^T,
\end{aligned}$$

with $\mathbf{H}_i(n)$ denoting the $L \times (2L-1)$ convolution matrix (also referred to as the Sylvester matrix) for the ith channel. By concatenating all M outputs of (2.3), a system of equations [122]

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{b}(n), \tag{2.4}$$

can be obtained where $\mathbf{x} = \left[\mathbf{x}_0^T\ \mathbf{x}_1^T\ \cdots\ \mathbf{x}_{M-1}^T\right]^T$, $\mathbf{H} = \left[\mathbf{H}_0^T\ \mathbf{H}_1^T\ \cdots\ \mathbf{H}_{M-1}^T\right]^T$, and $\mathbf{b} = \left[\mathbf{b}_0^T\ \mathbf{b}_1^T\ \cdots\ \mathbf{b}_{M-1}^T\right]^T$. The problem of Blind System Identification (BSI) is to find $\mathbf{h} = [\mathbf{h}_0^T\ \mathbf{h}_1^T\ \cdots\ \mathbf{h}_{M-1}^T]^T$ using only $\mathbf{x}$ up to a nonzero scale factor across all channels [1]. This scale factor is irrelevant in most acoustic signal processing applications. An acoustic enclosure (i.e. room) can generally be regarded as a stable system with the coefficients of $\mathbf{h}$ tending to zero with increasing tap number [2].

According to [123], two conditions are necessary and sufficient for blind identifiability of a SIMO system using BSI algorithms [1]:

1. *Channel diversity*: The channel transfer functions do not share any common zeros in their transfer functions [124], i.e., the polynomials formed by $\mathbf{h}_i$ $(0 \le i \le M)$ are co-prime;

2. *Full rank condition*: The autocorrelation matrix of the input signal $\mathbf{R}_{ss} = E\left\{\mathbf{s}(n)\mathbf{s}^T(n)\right\}$ is of full rank, where $E\{\cdot\}$ denotes the mathematical expectation operator, such that the SIMO system can be fully excited. This can be understood by first expressing, for the noiseless case [2],

$$\mathbf{S}(n)\mathbf{h}_i(n) = \mathbf{x}_i(n), \quad i = 0, 1, \cdots, M-1, \tag{2.5}$$

where $\mathbf{S}(n)$ is defined as the $L \times L$ Hankel matrix given by

$$\mathbf{S}(n) = \begin{bmatrix} s(n) & s(n-1) & \cdots & s(n-L+1) \\ s(n-1) & s(n-2) & \cdots & s(n-L) \\ \vdots & \vdots & \cdots & \vdots \\ s(n-L+1) & s(n-L) & \cdots & s(n-2L+2) \end{bmatrix}. \tag{2.6}$$

If $\mathbf{S}(n)$ is rank deficient, then (2.5) will not have a unique solution even if the source signal $\mathbf{s}(n)$ is known.

In this work we distinguish between supervised and unsupervised methods. Supervised system identification methods rely on either the system function or the input signal to be known while in unsupervised system identification methods the channel impulse responses $\mathbf{h}_i$ $(0 \leq i \leq M)$ are estimated from the observations $\mathbf{x}_i$ without utilizing the source signal $\mathbf{s}(n)$ [121]. Supervised methods are presented in the next section, while the interested reader should refer to [121, 125, 126] for an overview of blind adaptive algorithms in the context of estimating acoustic channels such as the Normalized Multichannel Frequency Domain Least Mean Squares (NMCFLMS) algorithm [127] and its robust implementation (RNMCFLMS) [128]. Other approaches to the BSI problem can be classified as least-squares methods [123, 129], subspace methods [130] and maximum-likelihood methods [131].

2.2 Room Acoustics

The science of room acoustics offers an understanding of the physical process by which sound waves propagate in enclosed spaces and the manner in which acoustic reflections combine to give the effect of what is referred to as reverberation [2]. This section aims to summarize some of the main concepts of room acoustics that are relevant to geometric room inference.

In principle, any complex sound field is considered as a superposition of plane waves. The propagation of such waves within a room can be considered to be a linear process after applying several simplifications including the assumptions that the medium in which the waves travel is homogenous, at rest, and that its characteristics are independent of the wave amplitude [7]. Then the propagation of acoustic waves through a material can be described by the second order partial differential wave equation. The

wave equation describes the evolution of sound pressure $p(\mathbf{q}, t)$, without any driving source, as a function of position $\mathbf{q} = (q_x, q_y, q_x)$ and time t and is given by [2, 132]

$$\nabla^2 p(\mathbf{q}, t) - \frac{1}{\mathfrak{c}^2} \frac{\partial^2 p(\mathbf{q}, t)}{\partial t^2} = 0, \tag{2.7}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \tag{2.8}$$

is the Laplacian expressed in the Cartesian coordinates (x, y, z), and $\mathfrak{c}$ is the speed of sound. The wave equation accurately describes the pressure in a realistic sound field provided that the wave amplitude is small such that $|p(\mathbf{q}, t)| \ll \rho_0 \mathfrak{c}^2$ where ρ_0 is the density of the propagation medium at equilibrium.

In order to calculate the sound field emanating from a source in a specific room we need an additional source function in (2.7) and boundary conditions that describe the sound reflection and absorption at the walls [133]. Let $s(\mathbf{q}, t)$ denote the source function, then the wave equation is given by

$$\nabla^2 p(\mathbf{q}, t) - \frac{1}{\mathfrak{c}^2} \frac{\partial^2 p(\mathbf{q}, t)}{\partial t^2} = -s(\mathbf{q}, t). \tag{2.9}$$

The wave equation can also be considered in the frequency domain. The Fourier transform of sound pressure $p(\mathbf{q}, t)$ is given by

$$P(\mathbf{q}; \omega) \triangleq \mathcal{F}\left\{p(\mathbf{q}, t)\right\}(\omega) = \int_{-\infty}^{\infty} p(\mathbf{q}, t) e^{-j\omega t}\, dt, \tag{2.10}$$

where $j = \sqrt{-1}$. The Fourier transform of (2.7) gives rise to the time-independent Helmholtz equation

$$\nabla^2 P(\mathbf{q}; \omega) + k^2 P(\mathbf{q}; \omega) = 0, \tag{2.11}$$

where

$$k = \frac{\omega}{\mathfrak{c}} = \frac{2\pi}{\lambda}$$

is the wavenumber, ω is the angular frequency and λ is the wavelength.

If there is a harmonic disturbance which is producing the waves, for which the source function is given by $s(\mathbf{q}, t) = S(\mathbf{q}; \omega) e^{j\omega t}$ then the Helmholtz equation is given by

$$\nabla^2 P(\mathbf{q}; \omega) + k^2 P(\mathbf{q}; \omega) = -S(\mathbf{q}; \omega). \tag{2.12}$$

It is often convenient to express the sound pressure at position $\mathbf{q}$, originating from a single source in a room, as the sum of the direct ($P_{\mathrm{d}}(\mathbf{q};\omega)$) and reverberant ($P_{\mathrm{r}}(\mathbf{q};\omega)$) components such that [7, 134]:

$$P(\mathbf{q};\omega) = P_{\mathrm{d}}(\mathbf{q};\omega) + P_{\mathrm{r}}(\mathbf{q};\omega). \tag{2.13}$$

Consider a plane wave-like approximation to the solution of (2.12) and write

$$P(\omega) = Ae^{-j\omega\tau}, \tag{2.14}$$

where A represents a slowly changing wave amplitude and $\omega\tau$ represent the wavefronts; analogously, the surfaces with constant τ are called timefronts (or phase function). By placing (2.14) into the form of (2.12), considering the high frequency approximation $\frac{\nabla^2 A}{A} \ll k^2$ and separating real and imaginary terms of the equation, one can obtain the *Eikonal* equation

$$(\nabla\tau)^2 = \frac{1}{\mathfrak{c}^2}, \tag{2.15}$$

and the *transport* equation

$$2\left(\nabla A \cdot \nabla\tau\right) + A\nabla^2\tau = 0. \tag{2.16}$$

There are different ways to solve the eikonal equation. In this work we consider the ray tracing approach. The limitation of this approach is that the phase function (τ) is given only along the ray, not on a uniform grid. Furthermore, the equation is fundamentally valid only in the limit $\omega \to \infty$. This implies that the eikonal equation (and many other ray-tracing techniques) may only be used when variations in velocity are negligible on spatial scales that are comparable to the wavelengths of the propagating waves.

2.2.1 Simulating Room Acoustics

The previous section describes, mathematically, the sound propagation in a room using the wave equation. Solving the wave equation (2.9) yields an impulse response from a source to a microphone. However, describing the impulse response in an analytical form is often not possible and consequently the solution must be approximated numerically. With reference to Figure 2.2, there are three classes of models that approximate the impulse responses: wave-based models, ray-based models and statistical models.

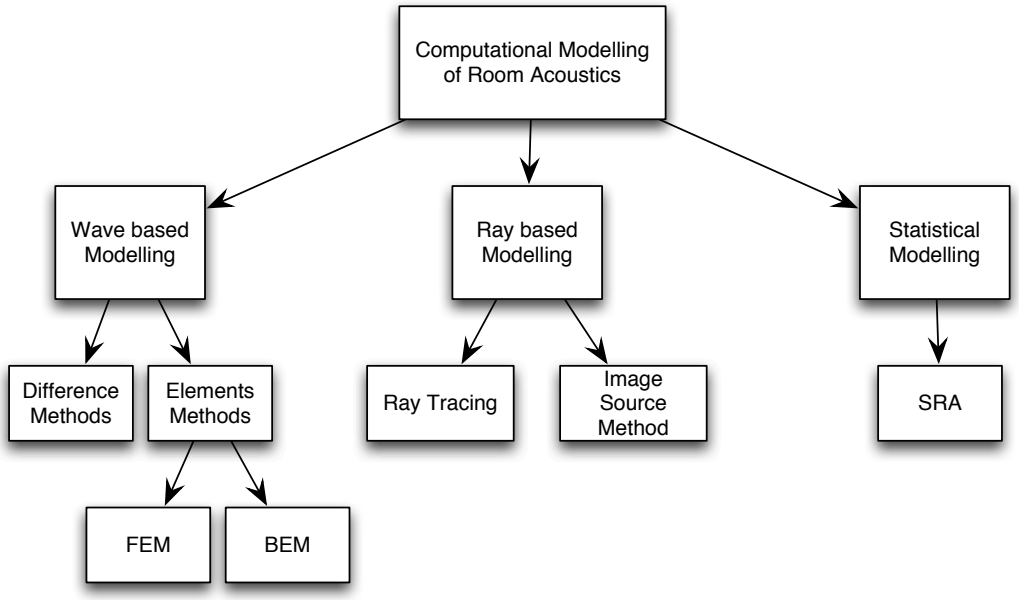


Figure 2.2: Methods for modeling and simulating room acoustics [1].

Wave-based methods, such as the Finite Element Method (FEM) [135], Boundary Element Method (BEM) [136], and Finite-Difference Time-Domain (FTDT) methods [137], are generally the most accurate at a penalty of high computational complexity. In wave-based methods it is usually highly challenging to incorporate appropriate boundary conditions and geometrical descriptions of the objects within the acoustic environment [2].

Ray-based methods are based on geometrical acoustics that can be simplified into a form in which sound waves are represented by rays and reflections are specular. In ray-tracing methods the sound power emitted by a sound source is described by a finite number of rays. These rays propagate through space and are reflected after every collision with the room boundaries [133]. It is advantageous to control the number of rays used in a simulation, which can be of the order 10^5 or more, by limiting the model to include only first and second order reflections in order to maintain the computational requirement of the simulation at a modest level [2]. A discussion into the accuracy of ray-tracing methods can be found in, for example [138].

Statistical Room Acoustics (SRA) considers amplitudes and phases of all reflected acoustic plane waves in a room as randomly distributed, forming a diffuse sound field

at any point in the room. Statistical modeling methods have been widely used in aerospace, ship and the automotive industry for high frequency noise analysis and acoustic designs. They are not suitable for auralization purposes since these methods do not model the temporal behaviour of a sound field [133].

In this work, ray-based methods are considered, such as the source-image method, originally proposed by Allen and Berkley [35]. It is one of the most commonly used techniques for simulating room acoustics in the context of speech dereverberation, sharing a closely related system model with the approach in this work. The image model can be used to simulate the impulse response in a room for a given source and microphone location. Usually, a single static sound source is considered and the SIMO approach obtained by computing FIRs that model the acoustic channels between a source and a receiver at different positions in a rectangular room. The image method was extended to arbitrary polyhedra with any number of sides in [36]. An illustration of the source-image method for a two-dimensional case is depicted in Figure 2.3, where the room is indicated with a bold rectangle. Note that in practice, the images extend over a three-dimensional lattice [2].

Image Method

Consider a rectangular room with length, width and height given by L_x, L_y, L_z. The sound source is located at $\mathbf{r}_\mathrm{s} \triangleq [x_\mathrm{s}\ y_\mathrm{s}\ z_\mathrm{s}]$ and the microphone at $\mathbf{r} \triangleq [x\ y\ z]$ with respect to the origin, i.e. [0 0 0], located at one of the corners of the room. The relative positions of the images measured with respect to the receiver position and obtained using the walls $x = 0$, $y = 0$ and $z = 0$ can be written as [133]

$$\mathbf{R_p} = [(1-2q)x_\mathrm{s} - x, (1-2j)y_\mathrm{s} - y, (1-2k)z_\mathrm{s} - z]. \tag{2.17}$$

Each of the elements in the triple $\mathbf{p} = (q, j, k)$ can take on values 0 or 1, resulting in eight different combinations that specify a set $\mathcal{P}$, i.e., $\mathcal{P} = \{(q, j, k) : q, j, k \in \{0, 1\}\}$. When the value of $\mathbf{p}$ is 1 in any dimension, then an image of the source in that direction is considered. Some of these images correspond to higher order reflections. In order to take these into account, we add the vector $\mathbf{R_m}$ to $\mathbf{R_p}$ where [133]

$$\mathbf{R_m} = [2m_xL_x, 2m_yL_y, 2m_zL_z], \tag{2.18}$$

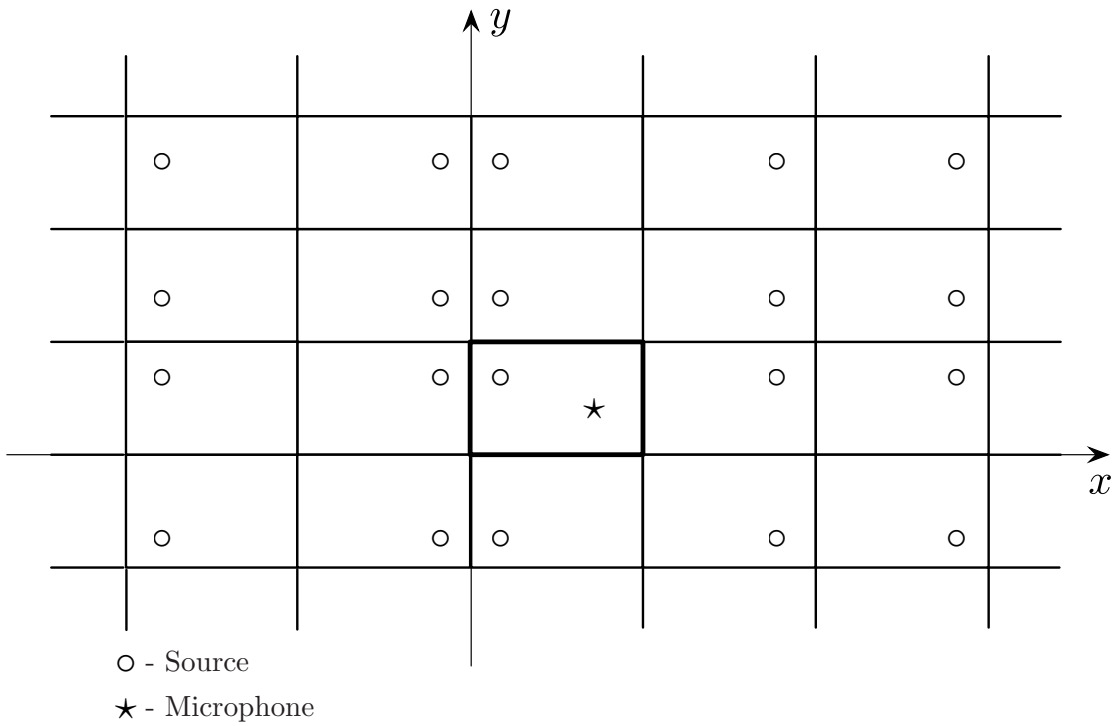


Figure 2.3: Example image sound sources for a rectangular room [2].

where m_x, m_y, and m_z are integer values. Each of the elements of the triple $\mathbf{m} = (m_x, m_y, m_z)$ takes on values from $-N$ to $+N$. The reflection order related to an image at position $\mathbf{R_p} + \mathbf{R_m} + \mathbf{r}$ is given by

$$O_{\mathbf{p},\mathbf{m}} = |2m_x - q| + |2m_y - j| + |2m_z - k| \,. \tag{2.19}$$

The distance d between any source image and the microphone can be evaluated as the Euclidean distance corresponding to

$$d = \|\mathbf{R_p} + \mathbf{R_m}\| \,. \tag{2.20}$$

The time delay of arrival of the reflected sound ray, corresponding to any source image, can be expressed as

$$\tau = \frac{d}{\mathfrak{c}} = \frac{\|\mathbf{R_p} + \mathbf{R_m}\|}{\mathfrak{c}}. \tag{2.21}$$

The impulse response for this source and microphone location can now be written as [133]

$$h(\mathbf{r}, \mathbf{r}_\mathrm{s}, t) = \sum_{\mathbf{p}\in\mathcal{P}} \sum_{\mathbf{m}\in\mathcal{M}} \beta_{x_1}^{|m_x - q|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y - j|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z - k|} \beta_{z_2}^{|m_z|} \frac{\delta(t-\tau)}{4\pi d}, \tag{2.22}$$

where $\mathcal{M} = \{(m_x, m_y, m_z) : -N \leq m_x, m_y, m_z \leq N\}$ denotes a set that contains all desired triples $\mathbf{m}$. The quantities β_{x_1}, β_{x_2}, β_{y_1}, β_{y_2}, β_{z_1}, and β_{z_2} are the reflection coefficients of the six walls. For a given N, this method computes $8(2N+1)^3$ different paths. Once the impulse response has been computed from (2.22), the source signal can be convolved with the impulse response to simulate the signal picked up by the microphone.

It is important to note that the delays given by (2.21) do not always fall at sampling instants. In a modification to the original method by Allen and Berkley, Peterson suggested in [139] that each impulse in (2.22) is replaced by the impulse response of a Hanning-windowed ideal low-pass filter of the form

$$\delta_{\text{LPF}}(t) = \begin{cases} 0.5\left(1 + \cos\left(\frac{2\pi t}{T_w}\right)\right)\text{sinc}(2\pi f_c t), & \text{for } -\frac{T_w}{2} < t < \frac{T_w}{2} \\ 0, & \text{otherwise} \end{cases}, \tag{2.23}$$

where T_w is the width (in time) of the impulse response, often taken as $T_w = \frac{4f_s}{1000}$, i.e. 4 ms, and f_c is the cut-off frequency of the low-pass filter, often taken as $f_c = f_s/2$. Each impulse $\delta(t-\tau)$ in (2.22) is first replaced by $\delta_{\text{LPF}}(t-\tau)$ and subsequently sampled. By doing this, true delays of arrival of the reflected signals are simulated accurately even at the original low sampling frequency [133].

2.2.2 Impulse Response Measurement Techniques

The impulse response of a linear system can be determined by exciting the system with a source signal, and cross-correlating the input and output. A RIR is defined as the time domain impulse response of a room from the acoustic source to the receivers [47]. It characterizes the effect of the sound emitted by the source reaching the receiver point on a direct path and after having been reflected and diffused by the room boundaries. Measurements of RIRs are mostly used in architectural and building acoustics, since they give access to the acoustical properties of a room. A second interest for highly accurate RIR measurements arises from its application in auralization, where recorded signals are convolved with a RIR. Measurements made for auralization are usually done using an artificial head as receiver, i.e. there are two receiving microphones and a pair of RIRs referred to as a binaural room impulse response (BRIR).

A comparison of supervised impulse response measurement techniques, using a Maximum Length Sequence (MLS) or Time-Stretched Pulse (TSP), is given in [42]. In the

following (with slight abuse of notation but keeping consistency with the referenced works) a quick overview of MLS and TSP is presented.

Maximum Length Sequence

Binary maximum-length sequences have certain properties that make them attractive as excitation signals for measurements of acoustical impulse responses. Several publications [43, 140] give detailed descriptions of the theory and the implementation of MLS measurement systems [47].

A MLS is a periodic sequence of binary numbers (± 1). The magnitude of a MLS is the same at all frequencies except at DC. Using binary sequences as measurement signals ensures the highest possible excitation level and therefore high noise immunity. An MLS is created by reading from a linear shift register, where defined register outputs are summed (modulo 2) and fed back to the shift register input. MLSs are pseudo-random noise sequences known as m-sequences in coding theory. They are generated by the primitive polynomials over Galois[1] fields of 2^m elements [47]. The coefficients of the primitive polynomials determine, which registers are combined in the feedback structure. The length of a MLS or rather its period is $L = 2^m - 1$ where m is the order of the sequence. There exists several primitive polynomials and thus several sequences of each order, while they are different in the sense of being cyclically distinct, i.e. they cannot be matched by linear rotation. The calculation of the primitive polynomials over Galois fields is not trivial especially for high orders [47]. In general, the linear transfer function of a system is calculated by a division in frequency domain,

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}. \tag{2.24}$$

In this way the excitation signal, X, is deconvolved from the system output, Y, to obtain the transfer function of the system, H. The division with X may cause problems if the magnitude response of the signal displays very low values at some frequencies. Hence, it is desirable to use an excitation signal that excites all frequencies equally much, such as an MLS. A sequence order m has to be chosen in advance depending on the expected length of the impulse response of the measured system. The length of the sequence L should be chosen sufficiently long to avoid time aliasing. In order to minimize the

[1] In abstract algebra, a finite field or Galois field (so named in honor of Évariste Galois) is a field that contains a finite number of elements.

amount of computation required by the cross-correlation step, the system can be excited by a binary maximal-length sequence, and the cross correlation performed using the fast Hadamard[1] transform. In this way, only additions are required, and the number of additions is linearithmic, i.e. 2.5 $m \log_2 m$ [43] compared to m^2 multiplications using straightforward cross-correlation techniques. Measuring the impulse response using a noise-like excitation is capable of providing much greater dynamic range than can be obtained using an impulsive excitation. When the noise-like excitation is chosen to be a binary maximal-length shift register sequence, several advantages accrue [43]. Furthermore, since measurements are exactly repeatable, additional improvement in the dynamic range can be achieved by averaging together the responses of several measurements.

Time-Stretched Pulse

This method is based on a time expansion and compression technique of an impulsive signal [45]. The aim of using an expansion process for the excitation signal is to increase the amount of sound power emitted for a fixed magnitude of the signal and therefore increase the signal-to-noise ratio without increasing the nonlinearities introduced by the measurement system [42].

The time-stretched pulse sequence was first proposed by Aoshima (ATSP) to measure transfer functions of acoustic systems, which according to [45] can be written as:

$$H(k) = \begin{cases} G(k)e^{ipk^2}, & \text{for } 0 \leq k < N/2 \\ G(N/2), & \text{for } k = N/2 \\ H(N-k), & \text{for } N/2 < k < N \end{cases}, \tag{2.25}$$

where $N = 2^m$ (m is an integer), $G(k)$ defines the power spectrum of the pulse, and p determines the stretch of the signal. The IDFT (inverse discrete Fourier transform) of $H(k)$, denoted $h(n)$, gives the time domain representation of the pulse. In order to use the pulse to measure the transfer function over the full frequency range from dc to $f_s/2$ (2.25) needs to be written as

$$H(k) = \begin{cases} e^{ipk^2}, & \text{for } 0 \leq k < N/2 \\ 1, & \text{for } k = N/2 \\ H(N-k), & \text{for } N/2 < k < N \end{cases}. \tag{2.26}$$

[1]The Hadamard transform is an example of a generalized class of Fourier transforms. It performs an orthogonal, symmetric, involutional, linear operation on 2^m real (or complex) numbers.

The absolute value of this function is constant 1 for all k. The IDFT of (2.26), i.e., $h(n)$ gives the signal in the time domain [45].

After the response of a system to this signal is measured, this stretched response is compressed again by using an inverse filter of $h(n)$, so that the impulse response of the system is obtained. The frequency domain representation $H^{-1}(k)$ of the inverse filter $h^{-1}(n)$ is defined as

$$H^{-1}(k) = \begin{cases} e^{-ipk^2}, & \text{for } 0 \leq k < N/2 \\ 1, & \text{for } k = N/2 \\ H \times (N-k), & \text{for } N/2 < k < N \end{cases} . \tag{2.27}$$

The steps to obtain the impulse response of the system are outlined as follows:

1. The DFT for the stretched response to the signal is calculated;
2. Then result of the DFT and the inverse $H^{-1}(k)$ from (2.27) are multiplied;
3. Finally, the IDFT for the products is calculated to obtain the desired impulse response.

Note, that the inverse filter in the time domain is obtained by reading the signal in reverse order, i.e. $h^{-1}(n) = h(N-n)$.

According to [42], there are certain advantages in choosing one method over the other, e.g. the TSP method with its particular timbre and high optimum output signal level, needed to mask out the ambient noise, make it unsuitable for occupied rooms. The drawback of the MLS method, on the other hand, lies in the tedious calibration that has to be carried out to obtain optimum results and in the appearance of spurious peaks (distortion peaks) due to the inherent nonlinearities of the measurement system.

2.3 Source Localization

Estimating the relative distance between sensors and source, i.e. the source range, along with the absolute position of the source (with respect to the receivers) is referred to as source localization in this work. The motivation for discussing both source range and position estimation is twofolds. Firstly, the range estimate, together with the TDOA measurements, enable the extraction of TOA information directly from unsynchronized

AIRs. Secondly, the source position estimate allows a fixed coordinate frame of reference, with respect to the system origin (which in this work coincides with the position of a designated reference sensor).

We consider two closed-form source position estimators under an additive measurement error model, which assumes that the errors are independent of the measurements. The first, [93], is a least-squares approach that is an approximation of the ML estimator when the TDOA estimation errors are small. It assumes knowledge of the TDOA covariance matrix which may not be known in practice. It is therefore a suitable starting point to highlight the effects of uncertainty and noise introduced to the TDOA measurements. The second estimator, [80], is also a least-squares approach, but makes no assumptions on the TDOA covariance matrix.

Both approaches use a two-step LS estimation procedure that exploits the independent measurements of both source range and position to impose a known constraint between the two to yield an improved position estimate. For ease of illustration we outline both approaches in a 2-D plane and as a consequence the original 3-D case outlined in [80] is adapted to 2-D.

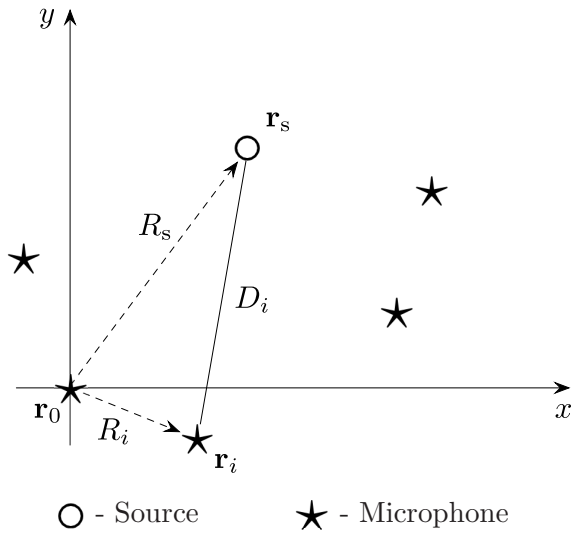


Figure 2.4: Localization in a 2-D plane.

2.3.1 Quadratic-Correction Least-Squares

With reference to Figure 2.4, assume that there are M microphones distributed arbitrarily in a 2-D plane located at

$$\mathbf{r}_i \triangleq [x_i \; y_i]^T, \quad i = 0, \cdots, M-1. \tag{2.28}$$

and a source located at $\mathbf{r}_{\rm s} \triangleq [x_{\rm s} \; y_{\rm s}]^T$. The first microphone ($i = 0$), commonly designating the closest microphone to the sound source, is placed at the origin of the coordinate system $\mathbf{r}_0 = [0 \; 0]^T$. The distances from the origin to all other microphones and source are given by their respective Euclidean norm such that

$$\begin{aligned} \|\mathbf{r}_i\| &= \sqrt{x_i^2 + y_i^2}, \quad i = 1, \cdots, M-1, \\ \|\mathbf{r}_{\rm s}\| &= \sqrt{x_{\rm s}^2 + y_{\rm s}^2}. \end{aligned}$$

The TDOAs from all possible receiver pairs are used to calculate the Gauss-Markov (weighted) estimate with respect to the first receiver [93], i.e. $\tau_{i,0}^{\rm GM} = \tau_i^{\rm GM} - \tau_0^{\rm GM}$ for $i = 1, 2, \cdots, M-1$. In other words, $\tau_{i,0}^{\rm GM}$ defines the TDOA of the direct-path coefficient for the ith channel with respect to the channel having the earliest direct-path. The TDOAs of the direct-path between microphones i and j are computed from

$$\tau_{i,j}^{\rm GM} = \left| \tau_{i,0}^{\rm GM} - \tau_{j,0}^{\rm GM} \right|, \quad i, j = 1, \cdots, M-1. \tag{2.29}$$

Let $\boldsymbol{\tau}^{\rm GM} = [\tau_{1,0}^{\rm GM}, \tau_{2,0}^{\rm GM}, \cdots, \tau_{M-1,0}^{\rm GM}]^T$ be the estimated TDOA vector. The covariance matrix $\boldsymbol{\Psi}$ of $\boldsymbol{\tau}^{\rm GM}$ is given by [93]

$$\begin{aligned} \boldsymbol{\Psi} = \Bigg\{ & \frac{2T}{2\pi} \int_0^{\Omega} \omega^2 \frac{S(\omega)^2}{1 + S(\omega) \mathrm{Tr}\left(\mathbf{N}\omega\right)^{-1})} \\ & \times \left[\mathrm{Tr}\left(\mathbf{N}(\omega)^{-1} \right) \mathbf{N}_p(\omega)^{-1} - \mathbf{N}_p(\omega)^{-1} \mathbf{1}\mathbf{1}^T \mathbf{N}_p(\omega)^{-1} \right] d\omega \Bigg\}^{-1}, \end{aligned} \tag{2.30}$$

where 0 to Ω is the frequency band processed and T is the observation time. $S(\omega)$ is the signal power spectrum, $\mathbf{N}(\omega) = \mathrm{diag}\{N_1(\omega), N_2(\omega), \cdots, N_M(\omega)\}$ is the noise power spectral matrix, $\mathbf{N}_p(\omega)$ is the lower right $M-1$ by $M-1$ partition of the matrix $\mathbf{N}(\omega)$ and $\mathbf{1}$ is a vector of unity which has the same size as $\mathbf{N}_p(\omega)$. The trace of the noise power spectral matrix, denoted $\mathrm{Tr}\,(\,\cdot\,)$, is equal to the sum of its diagonal elements. The squared distance between source and sensor i can be computed from

$$\begin{aligned} D_i^2 &= (x_i - x_s)^2 + (y_i - y_s)^2 \\ &= K_i - 2x_i x_s - 2y_i y_s + x_s^2 + y_s^2, \quad i = 0, \cdots, M-1, \end{aligned} \tag{2.31}$$

where we denote the squared norm of the microphone distances as

$$K_i \triangleq \|\mathbf{r}_i\|^2 = x_i^2 + y_i^2. \tag{2.32}$$

If $\mathfrak{c}$ is the signal propagation speed then

$$d_{i,0} = \mathfrak{c}\, \tau_{i,0}^{\mathrm{GM}} = |D_i - D_0|\,. \tag{2.33}$$

Based on this we can define a set of nonlinear equations whose solution gives (x_s, y_s). We denote R_s as the unknown source range (i.e. $R_\mathrm{s} = \|\mathbf{r}_\mathrm{s}\|$) and consequently define $\theta = [x_s\ y_s\ R_\mathrm{s}]^T$ as the unknown vector to be estimated. The elements of θ are related to (2.32) which means there is a set of nonlinear equations in two variables x and y. The approach to solve the nonlinear problem is to first assume that there is no relationship among x, y and R_s. They can then be solved by least squares [93]. Solving for both source location and range the generalized LS solution is then given by:

$$\theta = (\mathbf{G}_a^T \mathbf{\Psi}^{-1} \mathbf{G}_a)^{-1} \mathbf{G}_a^T \mathbf{\Psi}^{-1} \mathbf{h}, \tag{2.34}$$

$$\mathbf{h} = \frac{1}{2} \begin{bmatrix} d_{1,0}^2 - K_1 + K_0 \\ d_{2,0}^2 - K_2 + K_0 \\ \cdot \\ d_{M-1,0}^2 - K_{M-1} + K_0 \end{bmatrix}, \mathbf{G_a} = - \begin{bmatrix} x_{1,0} & y_{1,0} & d_{1,0} \\ x_{2,0} & y_{2,0} & d_{2,0} \\ \cdot & \cdot & \cdot \\ x_{M-1,0} & y_{M-1,0} & d_{M-1,0} \end{bmatrix},$$

where the symbols $x_{i,0}$ and $y_{i,0}$ stand for $x_i - x_0$ and $y_i - y_0$ respectively.

The proposed solution requires knowledge of the TDOA covariance matrix which may not be known in practice. Given the assumptions on the TDOA covariance matrix it is important to point out that this method yields an optimal estimator only when TDOA estimation errors are small. If the noise power spectral densities are similar at the sensors, the covariance matrix can be replaced by a matrix of diagonal elements σ_d^2 and $0.5\ \sigma_d^2$ for all other elements, where σ_d^2 is the TDOA variance. We define $\hat{\mathbf{\Psi}}$ as an approximation of the true covariance matrix of $\tau^{\mathbf{GE}}$. Solving for both source location and range we obtain the following approximation:

$$\hat{\theta} \approx (\mathbf{G}_a^T \hat{\mathbf{\Psi}}^{-1} \mathbf{G}_a)^{-1} \mathbf{G}_a^T \hat{\mathbf{\Psi}}^{-1} \mathbf{h}, \tag{2.35}$$

where

$$\hat{\mathbf{\Psi}} = \sigma_d^2 \begin{bmatrix} 1 & 0.5 & 0.5 & \cdots & 0.5 \\ 0.5 & 1 & 0.5 & \cdots & 0.5 \\ \cdot & \cdot & \cdot & \cdots & \\ 0.5 & 0.5 & 0.5 & \cdots & 1 \end{bmatrix}.$$

This two-step procedure, with quadratic correction, is in fact an approximation of a true ML estimator for emitter location. As a first step a weighted linear LS gives the initial solution. As a second step another weighted LS computation, using the known constraint between source coordinates and range, yields an improved position estimate. For a more exhaustive overview of the QCLS method please refer to [93].

2.3.2 Linear-Correction Least-Squares

The source localization algorithm outlined in [80] is based on a least squares estimator employing a spherical LS error criterion defined in a 3-D space. For our purposes this algorithm is modified to a 2-D space. Consequently, the spherical LS error function is simplified to a circular LS error criterion.

We again assume that there are M microphones distributed arbitrarily in a 2-D plane located at positions $\mathbf{r}_i$ with the reference microphone ($i = 0$) placed at the origin of the coordinate system, i.e., $\mathbf{r}_0 = [0\ 0]^T$, and a source located at $\mathbf{r}_\mathrm{s}$, as shown in Figure 2.4. The distances from the origin to the ith microphone and the source are denoted by R_i and R_s, respectively, where

$$R_i \triangleq \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2}, \quad i = 1, \ldots, M-1,$$
$$R_\mathrm{s} \triangleq \|\mathbf{r}_\mathrm{s}\| = \sqrt{x_\mathrm{s}^2 + y_\mathrm{s}^2}.$$

The difference in the distances of microphones i and j from the source is the *range difference*, $d_{i,j}$, and is proportional to the TDOA of the direct-path between the ith and jth microphone:

$$d_{i,j} = \mathfrak{c} \cdot \tau_{i,j}^{\mathrm{GM}}. \tag{2.36}$$

We observe that the correct source location should be at the intersection of a group of circles. The centre of each circle is equal to the location of the microphone and the radius of each circle is related to the source-microphone distance. Therefore, the best estimate of the source location will be the point that yields the shortest distance to those circles defined by the range differences and the hypothesized source range. From [80] we establish the distance D_i from the ith microphone to the source

$$\hat{D}_i = R_\mathrm{s} + \hat{d}_{i,0}, \tag{2.37}$$

where $\hat{(\cdot)}$ denotes an observation based on the measured range difference. The error function is then defined as the difference between the measured and true values, which

when writing them in a vector form gives

$$\mathbf{e}(\mathbf{r}_\mathrm{s}) = \mathbf{A}\theta - \mathbf{b}, \tag{2.38}$$

where

$$\mathbf{A} \triangleq \left[\mathbf{S} \mid \hat{\mathbf{d}}\right], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{M-1} & y_{M-1} \end{bmatrix},$$
$$\theta \triangleq \begin{bmatrix} x_\mathrm{s} \\ y_\mathrm{s} \\ R_\mathrm{s} \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{1,0}^2 \\ R_2^2 - \hat{d}_{2,0}^2 \\ \vdots \\ R_{M-1}^2 - \hat{d}_{M-1,0}^2 \end{bmatrix},$$

and $\mathbf{S} \mid \hat{\mathbf{d}}$ indicates that $\mathbf{S}$ and $\hat{\mathbf{d}}$ are stacked side-by-side with $\hat{\mathbf{d}} = [\hat{d}_{1,0}, \hat{d}_{2,0}, \ldots, \hat{d}_{M-1,0}]^T$. The corresponding LS criterion is then given by

$$J = \mathbf{e}^T\mathbf{e} = [\mathbf{A}\theta - \mathbf{b}]^T [\mathbf{A}\theta - \mathbf{b}]. \tag{2.39}$$

The solution for θ is given by [80]

$$\hat{\theta}_1 = \mathbf{A}^\dagger \mathbf{b}, \tag{2.40}$$

where $(\,\cdot\,)^\dagger$ denotes the pseudo-inverse. The estimate $\hat{\theta}_1$ is an unconstrained global least-squares minimizer of the circular error criterion. In the presence of measurement errors in the range differences, it deviates from its true value and can be expressed as

$$\hat{\theta}_1 = \theta + \Delta\theta. \tag{2.41}$$

Finding the LS solution based on the circular error criterion (2.39) is a linear minimization problem

$$\min_\theta (\mathbf{A}\theta - \mathbf{b})^T (\mathbf{A}\theta - \mathbf{b}), \tag{2.42}$$

subject to a quadratic constraint

$$\theta^T \mathbf{\Sigma} \theta = 0, \tag{2.43}$$

where $\mathbf{\Sigma} \triangleq \mathrm{diag}(1, 1, -1)$ is a diagonal and orthonormal matrix. Based on the circular error criterion (2.39), an additional correction based on Lagrange multipliers can be

used on $\hat{\theta}_1$ to generate a second-pass corrected estimate $\hat{\theta}_2$ of θ. Solving for θ the constrained least squares estimate [80] is obtained such that

$$\hat{\theta} = (\mathbf{A}^T\mathbf{A} + \lambda\boldsymbol{\Sigma})^{-1}\mathbf{A}^T\mathbf{b}, \tag{2.44}$$

where λ is yet to be determined. In order to find λ, we can impose the quadratic constraint directly by substituting (2.44) into (2.43), so that

$$\mathbf{b}^T\mathbf{A}(\mathbf{A}^T\mathbf{A} + \lambda\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}(\mathbf{A}^T\mathbf{A} + \lambda\boldsymbol{\Sigma})^{-1}\mathbf{A}^T\mathbf{b} = 0. \tag{2.45}$$

By using eigenvalue decomposition, the matrix $\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}$ can be diagonalized as

$$\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}, \tag{2.46}$$

where $\boldsymbol{\Lambda} = \text{diag}(\gamma_1, \cdots, \gamma_3)$ and $\gamma_i, i = 1, \cdots, 3$ are the eigenvalues of the matrix $\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}$. As shown in [80], by setting $\mathbf{p} = \mathbf{U}^T\boldsymbol{\Sigma}\mathbf{A}^T\mathbf{b}$ and $\mathbf{q} = \mathbf{U}^T\mathbf{A}^T\mathbf{b}$ we can write the function of the Lagrange multiplier as

$$f(\lambda) \triangleq \mathbf{p}^T(\boldsymbol{\Lambda} + \lambda\mathbf{I})^{-2}\mathbf{q}, \tag{2.47}$$

which is a polynomial of degree four[1]. Numerical root searching methods, such as the secant method, can be employed to obtain λ. Finally, using the first estimate of $\hat{\theta}_1$ from (5.8) we can produce a corrected estimate $\hat{\theta}_2$ using linear correction such that [80]

$$\hat{\theta}_2 = \left[\mathbf{I} + \lambda(\mathbf{A}^T\mathbf{A})^{-1}\boldsymbol{\Sigma}\right]^{-1}\hat{\theta}_1, \tag{2.48}$$

which is the final output.

2.4 Hough Transform

The Hough Transform (HT) is most commonly known as a feature extraction technique used in image analysis, computer vision, and digital image processing [141]. A common application of this technique is to detect geometrical shapes within a certain class by a voting procedure. This voting procedure is carried out in a parameter space, from which object candidates are obtained as local maxima in a so-called accumulator that is explicitly constructed by the algorithm for computing the HT.

[1] Accordingly in three dimensions it is a polynomial of degree six.

The HT considers the following normal parametrization [142]

$$\rho = x \cos\theta + y \sin\theta, \tag{2.49}$$

which specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. A point in the cartesian space maps to a sinusoid in the Hough parameter space that corresponds to all the lines passing through it. Conversely, points in the parameter space are transformed into lines in the Cartesian coordinate space. Given two points lying on a line with parameters (ρ, θ), in the Hough parameter space the sinusoids corresponding to these two points intersect at ρ, θ. Therefore, given point estimates (pixels) in the coordinate space, the parameters of a line corresponding to the best-fit of these estimates can be found. Figure 2.5 shows how the horizontal line that coincides with the positive x-axis (i.e. $\rho = 0$ and $\theta = 0$) is found from a collection of pixels. Let $\rho \in \mathbb{R}^+$ and $\theta \in [0, \pi]$. For each point $[x_j \; y_j]^T$ we calculate

$$\hat{\rho} = x_j \cos\hat{\theta} + y_j \sin\hat{\theta}. \tag{2.50}$$

The results are stored in an accumulator $\mathcal{A}$, initially set to zero, which is incremented at every step such that:

$$\mathcal{A}\left(\hat{\rho}, \hat{\theta}\right) = \mathcal{A}\left(\hat{\rho}, \hat{\theta}\right) + 1. \tag{2.51}$$

The position of the largest maximum of the accumulator given by

$$\left[\hat{\theta}_{\max}, \hat{\rho}_{\max}\right] = \arg\max\left\{\mathcal{A}\left(\rho, \theta\right)\right\}. \tag{2.52}$$

Most HT techniques employ certain techniques for estimating the orientation of feature points (edges) to restrict the ranges of values of θ a pixel may vote for. The estimation of the orientation of each edge pixel is often uncertain due to: 1) image noise (e.g., positional errors from quantization and sensor errors); 2) small neighborhood associated with the edge detection procedure and the inherent uncertainty with the procedure; 3) the parametric representation used to define a line [143].

Although not explicitly mentioned in this work, the HT can also be used for the detection of 3D objects in range data or 3D point clouds. The extension of classical HT for plane detection is quite straight forward. A plane is represented by its explicit equation $z = a_x x + a_y y + a_c$ for which we can use a 3D Hough space corresponding to a_x, a_y and a_c.

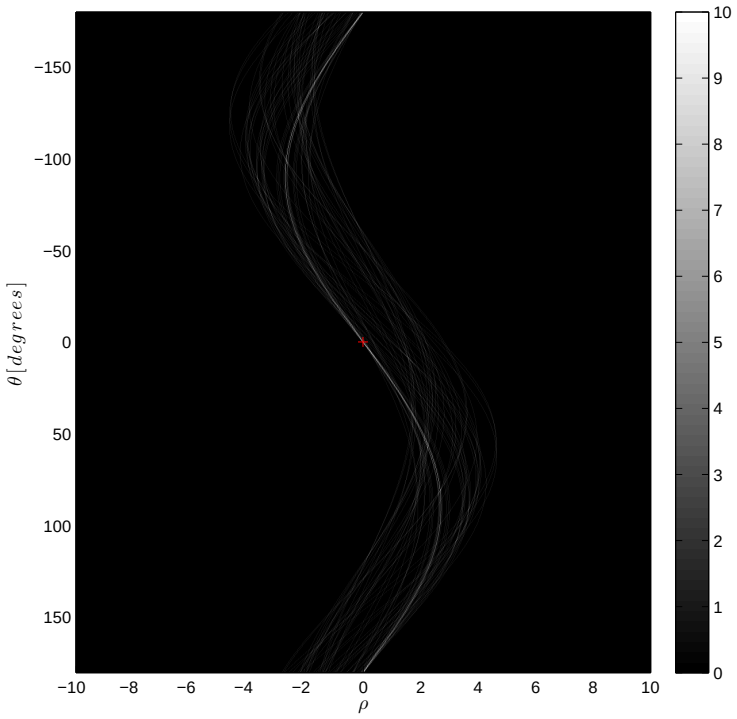


Figure 2.5: The Hough space graph showing the largest maximum (red '+').

In this work we consider the classical HT which was proposed by Duda and Hart [142] and describes the process of identification of lines in an image. The standard HT adopts a *top hat* strategy to compute the contribution of each point to a hypothesized line. Specifically, the scheme assumes all feature points located within a close range of the hypothesized line contribute equally to the line. The accumulator is, therefore, incremented by a unit for those feature points [143].

Conclusion

In this chapter a system model, based on the SIMO case, was presented that forms the foundation of the geometric inference framework. The mathematical formulation of waves traveling in an enclosed (and reverberant) space has been outlined. Solving the wave equation analytically is often impossible. A variety of numerical methods for simulation and actual impulse response measurement techniques were discussed. Two source localization methods, and their relevance to the system identification problem

were analysed. Finally, a method for estimating the parameters of a line segment, based on the Hough transform, was presented. The necessary background material given in this chapter, forms the foundation of the geometric inference framework that will be introduced in the subsequent chapters.

3

Geometric Localization Approach

Introduction

IN this chapter the geometric constraint, that permits the estimation of the line parameters of the reflectors in an acoustic scene, is introduced. The geometric constraint is developed in a two-dimensional plane, for illustration purposes.

The framework we consider makes use of time-of-arrival (TOA) and time-difference-of-arrival (TDOA) information, along with source and microphone positions, to form a set of elliptical constraints on the possible locations of the reflectors. The common tangent of these constraints is shown to correspond to the reflector location that can be found by minimizing a specific cost function in the least-squares sense.

We make no assumptions on how the TOA and TDOA information is obtained from the acoustic impulse responses (AIRs), and assume that the estimated AIRs, given by the vector $\hat{\mathbf{h}}_i(n)$, have been identified by either supervised or unsupervised techniques. We defer the analysis of the peak-picking problem from estimated AIRs to subsequent chapters.

Initially, the cost function for the single reflector case is developed and later extended to the case of multiple reflectors. We also highlight the permutation problem in the TOA information matrix.

The remainder of this chapter is organized as follows: Section 3.1 formulates the problem and the notation used in this work. Section 3.2 introduces the geometric constraint used in the localization of line reflectors for the 2-D case. Section 3.3 outlines

the common tangent estimation algorithm and extends the analysis and notation to the case where multiple reflectors are present in the acoustic environment.

Relevant Publications

1. J. Filos, E. A. P. Habets and P. A. Naylor, "A Two-Step Approach To Blindly Infer Room Geometries," *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, August 30 - September 2, 2010.

2. F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor and S. Tubaro, "Inference of Room Geometry from Acoustic Impulse Responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 26832695, Dec. 2012.

3.1 Problem Formulation

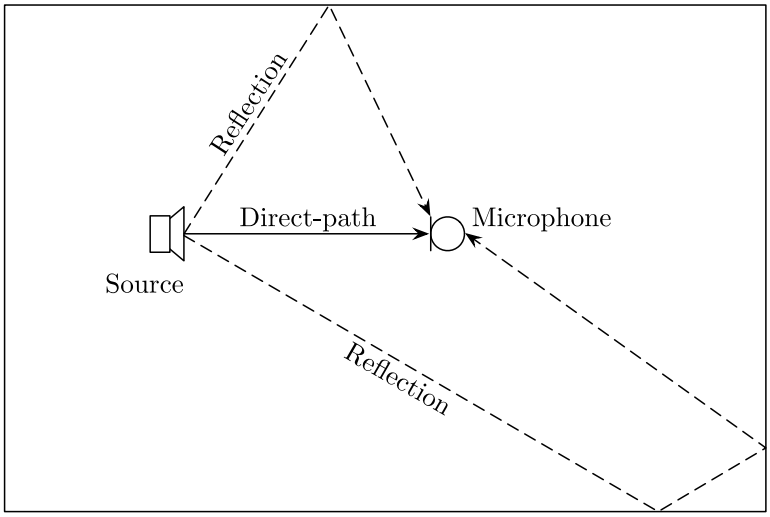


Figure 3.1: Schematic illustration of first and second order reflections in a room.

When acoustic signals are obtained in an enclosed space by one or more microphones positioned at a distance from the source, the observed signal consists of a superposition of many delayed and attenuated copies of the signal due to multiple reflections from the surrounding walls and other objects, as illustrated in Figure 3.1. We define the direct-path as the acoustic propagation path from the source to the microphone without

reflections. We also note that a delay of the superimposed copies arises because all other propagation paths are longer than the direct-path and that additional attenuation occurs at each reflection due to frequency dependent absorption [2].

With reference to Figure 3.2(a), a sound source located at $\mathbf{r}_\mathrm{s} \triangleq [x_\mathrm{s}\ y_\mathrm{s}]^T$ emits the signal $s(t)$. The signals $x_i(t)$ are observed by M microphones at positions

$$\mathbf{r}_i \triangleq [x_i\ y_i]^T, \quad i = 0, \ldots, M-1. \tag{3.1}$$

3.1.1 Acoustic Impulse Responses

The Acoustic Impulse Response (AIR) characterizes the acoustics of a given enclosure. This section will introduce the mathematical representation of the AIRs and discuss specific characteristics and assumptions made. Note, that the term AIR is used to refer to acoustic impulse responses in general, there are some cases where it is more appropriate to limit the acoustic context to be within a room, in which case, the impulse response is referred to as a Room Impulse Response (RIR). In this work, both terms are used interchangeably.

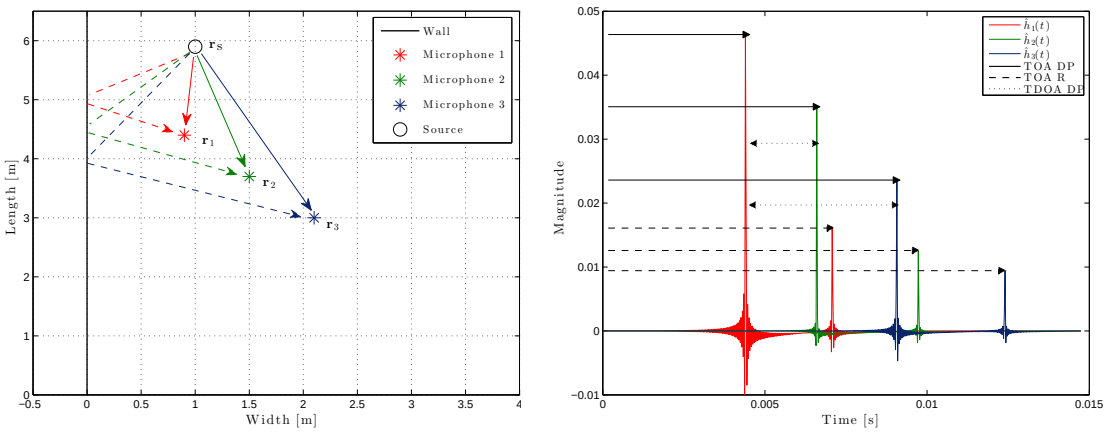
The observed signals, i.e. $x_i(t)$, are given by the convolution of the source $s(t)$ with the corresponding acoustic room impulse responses $h_i(t),\ i = 0, \ldots, M-1$:

$$x_i(t) = \int_0^\infty h_i(t')s(t-t')\,dt' + b_i(t),\ i = 0, \ldots, M-1, \tag{3.2}$$

where $b_i(t)$ is additive environmental noise assumed to be zero-mean and uncorrelated with the source signal. Under the hypothesis of ideal reflections the AIRs are given by

$$h_i(t) = \sum_{q=0}^{Q} \alpha_{i,q}\delta(t-\tau_{i,q}), \tag{3.3}$$

where Q is the total number of reflections of all orders, $\alpha_{i,q}$ is an attenuation term and $\tau_{i,q}$ is defined as the TOA associated with the ith microphone and the qth reflection. Note that the TOA of the direct-path is defined with respect to the *null* reflector, i.e. $q = 0$. With reference to Fig. 3.2(b) we can obtain estimates of the TOAs related to the direct path and reflections by analyzing $h_i(t)$. For this we note that the first peak in $h_i(t)$ is related to the time of propagation of the direct-path from $\mathbf{r}_\mathrm{s}$ to $\mathbf{r}_i$ given by $\tau_{i,0}$. Any subsequent peak in $h_i(t)$ is related to the composite time of propagation of the sound due to reflection. By defining $\mathbf{r}_{\mathrm{p},i}$ as the reflection point on any reflector, we obtain $\tau_{i,k},\ k = 1, \ldots, K$ as the sum of the propagation times from $\mathbf{r}_\mathrm{s}$ to $\mathbf{r}_{\mathrm{p},i}$, and then from $\mathbf{r}_{\mathrm{p},i}$ to $\mathbf{r}_i$ for any K reflectors present in our environment.



(a) Geometry of the problem: A single reflector in an anechoic space. Three microphones pick up the direct-path (solid line) along with the reflection (dashed line).

(b) AIR of three microphone estimates. The solid line TOA DP refers to the time of flight of the direct-path, the dashed line TOA R refers to the composite time of flight from source to reflector and then from reflector to microphone. TDOA DP refers to the time difference of arrival between the TOAs of the direct-path between microphone 1 and the other two.

Figure 3.2: Problem geometry and associated AIRs.

3.1.2 System Model

Let $s(n)$, $h_i(n)$, $x_i(n)$ represent sampled versions of the source, channel and observation at microphone i respectively. From the system equation in (2.3), i.e. $\mathbf{x}_i(n) = \mathbf{H}_i(n)\mathbf{s}(n) + \mathbf{b}_i(n)$, we assume that the estimated AIRs, given by the vector $\hat{\mathbf{h}}_i(n)$, have been identified by either supervised or unsupervised techniques. Peaks within the AIR correspond to the direct-path from source to receiver and the summed reflective paths from source to reflector and reflector to receiver. We assume that we have access to the temporal location of these peaks, which are found as discrete times $n_{i,k}$, with $n_{i,k} = \tau_{i,k} f_s$ in the case of synchronized measurements. For illustration purposes, we restrict ourselves to first-order reflections, when in reality, in closed geometries, infinitely many higher-order reflections occur.

We assume that in $h_i(n)$ the first echo after the direct-path is related to the same reflector. This assumption is satisfied if i) $\mathbf{r}_s$ is sufficiently close to the reflector of interest and ii) microphones are compactly organized in space. If both conditions

apply, then

$$h_i(n) = \alpha_{i,0}\delta(n - n_{i,0}) + \alpha_{i,1}\delta(n - n_{i,1}) + \sum_{k=2}^{N} \alpha_{i,k}\delta(n - n_{i,k}) \,, \tag{3.4}$$

where $\alpha_{i,k}$ is the attenuation along the direct ($k = 0$) or reflective ($k \geq 1$) path. Moreover, $n_{i,0} < n_{i,1} < \ldots < n_{i,k}^{(k)}$ for $k \geq 2$. The direct and the shortest reflective paths have been kept out the summation, as they constitute the input of the common tangent algorithm introduced in this work.

Consequently, we define the $M \times (K+1)$ matrix $\boldsymbol{\tau}$ containing TOA information for K reflectors such that

$$\boldsymbol{\tau} \triangleq \left[\boldsymbol{\tau}^{\mathrm{DP}} \mid \boldsymbol{\tau}^{\mathrm{RE}}\right], \tag{3.5}$$

where

$$\boldsymbol{\tau}^{\mathrm{DP}} \triangleq \begin{bmatrix} \tau_{0,0} \\ \tau_{1,0} \\ \vdots \\ \tau_{M-1,0} \end{bmatrix}, \quad \boldsymbol{\tau}^{\mathrm{RE}} \triangleq \begin{bmatrix} \tau_{0,1} & \tau_{0,2} & \cdots & \tau_{0,K} \\ \tau_{1,1} & \tau_{1,2} & \cdots & \tau_{1,K} \\ \vdots & \vdots & \cdots & \vdots \\ \tau_{M-1,1} & \tau_{M-1,2} & \cdots & \tau_{M-1,K} \end{bmatrix}.$$

The $M \times 1$ column vector $\boldsymbol{\tau}^{\mathrm{DP}}$ represents the direct-path TOAs, and is stacked side-by-side with the $M \times K$ matrix $\boldsymbol{\tau}^{\mathrm{RE}}$, that contains all first-order reflective path TOAs.

Additionally we define $\tau_{i,j}^{\mathrm{TDOA}}$ as the TDOA of the direct-path between the ith and jth microphone:

$$\tau_{i,j}^{\mathrm{TDOA}} = \left|\tau_{i,0} - \tau_{j,0}\right| . \tag{3.6}$$

Single Reflector Case

We shall assume in the following, without loss of generality, that only a single reflector (i.e. $K = 1$) is present in an otherwise anechoic environment. We make this assumption in order to outline the reflector localization procedure based on the geometric constraint, for ease of illustration and to point out an important challenge when estimating $\boldsymbol{\tau}$, and more specifically $\boldsymbol{\tau}^{\mathrm{RE}}$. Defining TOAs related to the direct-path and reflective-path propagation separately, for the single reflector case we temporarily drop the index k from $\tau_{i,k}$ such that $\tau_i^{\mathrm{DP}}, i = 0, \cdots, M-1$ and $\tau_i^{\mathrm{RE}}, i = 0, \cdots, M-1$ to denote the M direct and reflective-path TOAs respectively. Evidently, $\tau_i^{\mathrm{DP}} = [\tau_{0,0}\ \tau_{1,0}\ \cdots\ \tau_{M-1,0}]^T = \boldsymbol{\tau}^{\mathrm{DP}}$, i.e. the direct-path TOA vector remains unchanged. On the other hand, τ_i^{RE} is now a $M \times 1$ column vector chosen from the $M \times K$ matrix $\boldsymbol{\tau}^{\mathrm{RE}}$ such that all elements in

the τ_i^{RE} correspond to TOAs estimated for a single and common reflector. It becomes clear, that in the case where multiple reflectors are considered, the elements in $\boldsymbol{\tau}^{\mathrm{RE}}$ need to be ordered in a column wise fashion, such that each column represents TOA information related to the same reflector. Let π be the permutation operator that acts on matrices, in such a way that if $\boldsymbol{\tau}^{\mathrm{RE}}$ is some matrix related to a room with K reflectors, then $\pi(\boldsymbol{\tau}^{\mathrm{RE}})$ is the matrix that we would get if we permute the elements according to π. Consequently

$$\boldsymbol{\Gamma}_{\pi} = \pi(\boldsymbol{\tau}^{\mathrm{RE}}), \tag{3.7}$$

denotes the correctly permuted matrix of $\boldsymbol{\tau}^{\mathrm{RE}}$ such that each column contains TOA information for the same reflector. For the case $K = 1$, the permutation problem reduces to the trivial form $\boldsymbol{\tau}^{\mathrm{RE}} = [\tau_{0,1}\ \tau_{1,1}\ \cdots\ \tau_{M-1,1}]^T = \tau_i^{\mathrm{RE}}, i = 0, \cdots, M-1$.

3.2 Geometric Constraint

In this work we consider sound waves that can be represented by rays and reflections are therefore assumed specular. Under these assumptions, the relationship between

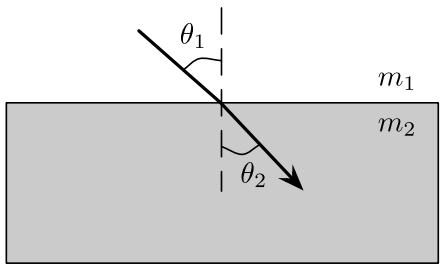


Figure 3.3: Relationship between angles of incidence and refraction of a wave according to Snell's law.

angles of incidence and refraction for a wave impinging on an interface between two media with different indices of refraction, i.e. m_1 and m_2, are given by Snell's law, as depicted schematically in Figure 3.3. The law follows from the boundary condition that a wave be continuous across a boundary, which requires that the phase of the wave be constant on any given plane, resulting in

$$m_1 \sin\theta_1 = m_2 \sin\theta_2, \tag{3.8}$$

where θ_1 and θ_2 are the angles from the normal of the incident and refracted waves respectively.

We make the fundamental assumption that source and receivers lie on the same plane and the lying plane of the reflector is orthogonal to this plane. In this scenario, the geometry of the acoustic scene is described by the plane in which sources and receivers lie. In a two-dimensional geometry a line is the collection of points $[x\ y]^T$ such that

$$\mathcal{L} = \left\{(x, y) \in \mathbb{R}^2 | l_1 x + l_2 y + l_3 = 0\right\}, \tag{3.9}$$

which after setting the line parameter $\mathbf{l} = [l_1\ l_2\ l_3]^T$ can be written as

$$\mathbf{l}^T [x\ y\ 1] = 0. \tag{3.10}$$

We observe that τ_i^{RE} is the sum of two terms: the propagation time from the source $\mathbf{r}_\mathrm{s}$ to the unknown reflection point $\mathbf{r}_{\mathrm{p},i}$ on the reflector $\mathbf{l}$ and the propagation time $\mathbf{r}_{\mathrm{p},i}$ to the microphone position $\mathbf{r}_0$. The knowledge of τ_i^{RE}, $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$, therefore, bounds the reflection point $\mathbf{r}_{\mathrm{p},i}$ to lie on an ellipse with foci in $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$ and whose major diameter is $r_i = \mathfrak{c}\tau_i^{\mathrm{RE}}$. We assume, moreover, that the reflection undergoes Snell's law, therefore the line perpendicular to $\mathbf{l}$ is also the bisector of the angle $\widehat{\mathbf{r}_\mathrm{s}\mathbf{r}_{\mathrm{p},i}\mathbf{r}_i}$, as shown in Figure 3.5. According to the properties of ellipses, this means that $\mathbf{l}$ is tangential to the ellipse. We notice that the reflection point $\mathbf{r}_{\mathrm{p},i}$ depends on the positions $\mathbf{r}_i$ and $\mathbf{r}_\mathrm{s}$, therefore if

Figure 3.4: The TOA of the reflective path is constituted by the time of propagation from $\mathbf{r}_\mathrm{s}$ to $\mathbf{r}_{\mathrm{p},i}$ and from $\mathbf{r}_{\mathrm{p},i}$ to $\mathbf{r}_i$. Possible reflection points lie on an ellipse.

we consider another microphone in the set $\mathbf{r}_i, i = 0, \ldots, M-1$, the reflection point on the ellipse changes. However, what remains unchanged for all the ellipses is that they are all tangential to the reflector. The common tangent estimation takes inspiration from the observation: the reflector line is found as the line that is tangential to all M ellipses. In order to accomplish this task, however, we need to parameterize the above tangential constraint in such a way that the tangent lines directly appear in the ellipse equation. This is where the projective geometry provides a convenient formulation.

3.2.1 Ellipse Parametrization

The aim of this section is to find the parameters of the ellipse given the foci $\mathbf{r}_i = [x_i\ y_i]^T$ and $\mathbf{r}_\mathrm{s} = [x_\mathrm{s}\ y_\mathrm{s}]^T$ and the major axis D_i. Using the parameters $\{a, b, c, d, e, f\}$ the conic can be expressed as [144]

$$\mathcal{C} = \left\{(x, y) \in \mathbb{R}^2 | ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0\right\}. \tag{3.11}$$

A parametrization that is convenient for our purposes is based on the representation of points using homogeneous coordinates. The homogeneous coordinates for the point $[x\ y]^T$ are $[\lambda x\ \lambda y\ \lambda]^T$, λ being a scalar different from zero. In such a representation the point $[x\ y]^T$ in the Euclidean space is mapped into a three-dimensional space and all points aligned on the direction $[\lambda x\ \lambda y\ \lambda]^T$ correspond to the same point in the Euclidean space, thus defining an equivalence class between homogeneous and Euclidean coordinates.

In homogeneous coordinates, the conic in (3.11) becomes

$$\mathbf{x}^T\mathbf{C}\mathbf{x} = 0\,, \tag{3.12}$$

where $\mathbf{x} = [\lambda x\ \lambda y\ \lambda]^T$ and $\mathbf{C}$ is the conic matrix, using the coefficients from (3.11), given by

$$\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}. \tag{3.13}$$

This defines an ellipse after constraining

$$\det(\mathbf{C}) \neq 0, \quad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \quad \det(\mathbf{C})/(a + c) < 0. \tag{3.14}$$

Assigning values to the parameters a,b,c,d,e,f, can be done in two ways. The first approach is to decompose the matrix in (3.13) as a product of translation, rotation and scaling matrices. The second approach is by expanding the implicit equation of the ellipse and comparing term by term with the parameters.

3.2.1.1 Matrix Decomposition

In order to justify a matrix decomposition of the conic in (3.13), we first note that the points on a unit circle satisfy

$$\mathbf{x}^T\mathbf{C}_\mathrm{I}\mathbf{x} = 0, \tag{3.15}$$

where

$$\mathbf{C}_\mathrm{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

We observe that for matrix $\mathbf{C}_\mathrm{I}$ the positive index of inertia is equal to two and that the negative index of inertia is equal to one [145]. In other words $\mathbf{C}_\mathrm{I}$ has one negative and two positive eigenvalues. We will show that this is the same for the matrix $\mathbf{C}$.

By Sylvester's law of inertia [145] we state that the negative index of inertia for $\mathbf{C}$ is equal to the number of sign changes in the sequence

$$\triangle_0 = 1, \quad \triangle_1 = a, \quad \triangle_2 = \begin{vmatrix} a & b \\ b & c \end{vmatrix}, \quad \triangle_3 = \det(\mathbf{C}).$$

Assuming $a > 0$ and therefore $c > 0$ we obtain $\triangle_1 > 1$ and since $ac - b^2 > 0$ it follows that $\triangle_2 > 0$ because $ac > 0$ and therefore $\operatorname{sgn}\{a\} = \operatorname{sgn}\{b\}$, where $\operatorname{sgn}\{\,\cdot\,\}$ is defined as

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0\,, \\ 0 & \text{if } x = 0\,, \\ 1 & \text{if } x > 0\,. \end{cases}$$

From $\det(\mathbf{C})/(a + c) < 0$ it follows that $\triangle_3 < 0$. The sequence $\triangle_0 > 0$, $\triangle_1 > 0$, $\triangle_2 > 0$, $\triangle_3 < 0$ has one sign change and therefore the positive index of inertia is equal to two and the negative index of inertia is equal to one.

Proposition 1. *Any matrix* $\mathbf{C}$*, describing an ellipse, i.e. satisfying conditions* (3.14)*, can be represented as a translated, rotated and scaled version of a unit circle given by* $\operatorname{diag}\{1, 1, -1\}$.

Proof. Because $\mathbf{C}$ is symmetric we can write it in the following form

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T,$$

with $\mathbf{\Lambda} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{I}_\mathrm{c}\mathbf{\Lambda}^{\frac{1}{2}}$, where $\mathbf{\Lambda}^{\frac{1}{2}}$ is defined as $|\mathbf{\Lambda}|^{\frac{1}{2}}$ and $\mathbf{I}_\mathrm{c} = \operatorname{diag}\{1, 1, -1\}$. From this it follows that

$$\mathbf{C} = (\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}})\mathbf{I}_\mathrm{c}(\mathbf{U}^T\mathbf{\Lambda}^{\frac{1}{2}}),$$

and by setting $\mathbf{\Phi} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$ we can justify the decomposition

$$\mathbf{C} = \mathbf{\Phi}\mathbf{I}_\mathrm{c}\mathbf{\Phi}^T.$$

Substituting into (3.15) and noting that $\mathbf{I}_\mathrm{c} = \mathbf{C}_\mathrm{I}$ this becomes

$$\mathbf{x}^T\mathbf{\Phi}\mathbf{C}_\mathrm{I}\mathbf{\Phi}^T\mathbf{x} = 0.$$

Let $\mathbf{T}$ be a translation matrix, then

$$\mathbf{x}^T\mathbf{T}^T(\mathbf{T}^{-1})^T\mathbf{\Phi}\mathbf{C}_\mathrm{I}\mathbf{\Phi}^T\mathbf{T}^{-1}\mathbf{T}\mathbf{x} = 0.$$

Setting $\mathbf{w} = \mathbf{T}\mathbf{x}$ it follows that

$$\mathbf{w}^T(\mathbf{T}^{-1})^T\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{C}_\mathrm{I}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T\mathbf{T}^{-1}\mathbf{w} = 0.$$

and finally by $\mathbf{z} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T\mathbf{T}^{-1}\mathbf{w} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T\mathbf{x}$ it follows that

$$\mathbf{z}^T\mathbf{C}_\mathrm{I}\mathbf{z} = 0.$$

This means that the first two components of $\mathbf{z}$ belong to the unit circle. In other words if (x, y) belongs to the ellipse, then there is a unique $(\tilde{x}, \tilde{y})$ that belongs to the unit circle with $\tilde{x} = (\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\mathbf{x})_1$ and $\tilde{y} = (\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}\mathbf{x})_2$. □

This justifies the following decomposition for the ellipse associated with the ith microphone and a single common reflector

$$\mathbf{C}_i = \mathbf{T}_i^{-T}\mathbf{R}_i^{-T}\mathbf{S}_i^{-T}\mathbf{C}_\mathrm{I}\mathbf{S}_i^{-1}\mathbf{R}_i^{-1}\mathbf{T}_i^{-1}, \tag{3.16}$$

where we can define a translation, rotation and scaling matrix such that

$$\mathbf{T}_i = \begin{bmatrix} 1 & 0 & \Delta x_i \\ 0 & 1 & \Delta y_i \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.17}$$

$$\mathbf{R}_i = \begin{bmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.18}$$

$$\mathbf{S}_i = \begin{bmatrix} Q_i^{\mathrm{smaj}} & 0 & 0 \\ 0 & Q_i^{\mathrm{smin}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.19}$$

The quantities Δx_i, Δy_i, γ_i, Q_i^{smaj} and Q_i^{smin} are defined as follows. The point at $(\Delta x_i, \Delta y_i)$ can be seen as the geographic midpoint between $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$ and is defined by

$$\begin{aligned} \Delta x_i &\triangleq x_\mathrm{s} + \frac{D_i\cos(\gamma_i)}{2}; \\ \Delta y_i &\triangleq y_\mathrm{s} + \frac{D_i\sin(\gamma_i)}{2}, \end{aligned}$$

with $\gamma_i \triangleq \tan^{-1}\left(\frac{y_s - y_i}{x_s - x_i}\right)$. The scaling of the semi-major and semi-minor axes of each ellipse is then given by

$$\begin{aligned} Q_i^{\text{smaj}} &= \frac{\mathfrak{c} \cdot \tau_i^{\text{RE}}}{2}; \\ Q_i^{\text{smin}} &= \frac{\sqrt{\left(\mathfrak{c} \cdot \tau_i^{\text{RE}}\right)^2 - D_i^2}}{2}, \end{aligned}$$

respectively.

3.2.1.2 Implicit Solution

The implicit equation of an ellipse with foci in (x_s, y_s) and (x_i, y_i) and with major diameter $r_i = \mathfrak{c}\tau_i^{\text{RE}}$ is

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} + \sqrt{(x - x_s)^2 + (y - y_s)^2} = r_i \ . \tag{3.20}$$

By expanding (3.20) and comparing term by term with a conic with parameters $(a_i, b_i, c_i, d_i, e_i, f_i)$ we obtain

$$\begin{aligned} a_i &= r_i^2 - (x_s - x_i)^2 \ , \\ b_i &= -2(x_s - x_i)(y_s - y_i) \ , \\ c_i &= r_i^2 - (y_s - y_i)^2 \ , \\ d_i &= (x_i - x_s)(x_i^2 + y_i^2 - x_s^2 - y_s^2 - r_i^2) - x_s r_i^2 \ , \\ e_i &= (y_i - y_s)(x_i^2 + y_i^2 - x_s^2 - y_s^2 - r_i^2) - y_s r_i^2 \ , \\ f_i &= (x_s^2 r_i^2 + y_s^2 r_i^2 \frac{y_i^2 - y_s^2 + x_i^2 - x_s^2 - r_i^2}{4} \ . \end{aligned}$$

More meaningful for our purposes is the definition of the *line conic* associated with the point conic defined in (3.12) and (3.13). The line $\mathbf{l}$ passes through the point $\mathbf{x}$ iff $\mathbf{l}^T\mathbf{x} = 0$ and is tangential to the point ellipse $\mathbf{C}_i$ iff [144]

$$\mathbf{l}^T \mathbf{C}_i^* \mathbf{l} = 0, \tag{3.21}$$

where $\mathbf{C}_i^* = \det(\mathbf{C}_i)\mathbf{C}_i^{-1}$ is the adjoint of the conic matrix $\mathbf{C}_i$. This formulation is important for the common tangent estimation algorithm that is presented in the next section.

3.3 Common Tangent Estimation Algorithm

As mentioned previously, if we acquire multiple impulse responses from $\mathbf{r}_\mathrm{s}$ to $\mathbf{r}_i, i = 0, \ldots, M-1$, the line $\mathbf{l}$ is tangent to all the ellipses $\mathbf{C}_i, i = 0, \ldots, M-1$ at points $\mathbf{r}_{\mathrm{p},i}, i = 0, \ldots, M-1$. Figure 3.5 shows an example where $M = 3$. Combining the

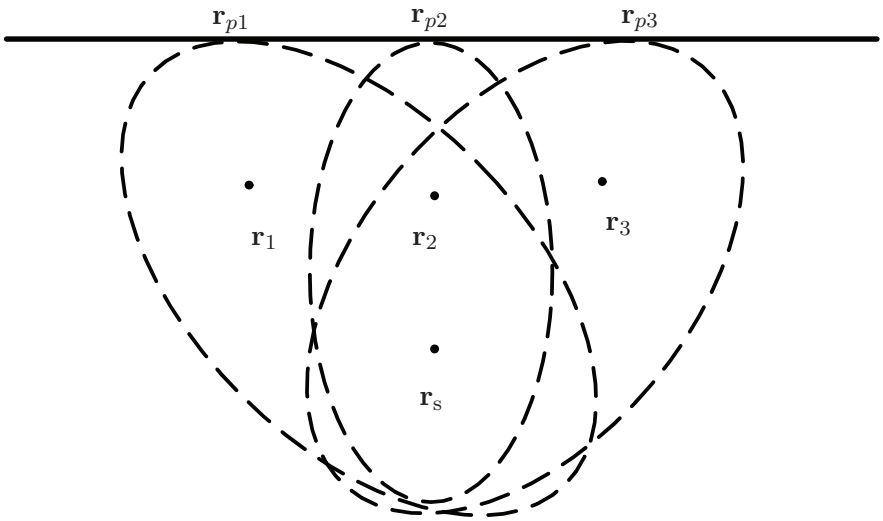


Figure 3.5: The reflector line is the common tangent to the ellipses traced for $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$.

constraints in (3.21), the reflector line is the simultaneous solution of [113, 114]

$$\begin{cases} \mathbf{l}^T\mathbf{C}_0^*\mathbf{l} & = 0 \\ \mathbf{l}^T\mathbf{C}_1^*\mathbf{l} & = 0 \\ \cdots & \\ \mathbf{l}^T\mathbf{C}_{M-1}^*\mathbf{l} & = 0 \end{cases} . \tag{3.22}$$

Since we have three unknowns (the parameters l_1, l_2, l_3) we need at least $M = 3$.

From a geometrical standpoint, solving (3.22) corresponds to finding the line $\mathbf{l}$, in the line parameter space, that lies on all the manifolds representing constraints in (3.22).

The solution of a nonlinear system as in (3.22) is difficult when the measures of τ_i^{DP} are affected by measurement errors and the positions $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$ are known only up to some uncertainty. We shall combine the individual equations in (3.22) into the cost function [113, 114]

$$J\left(\mathbf{l}, \{\mathbf{C}_i\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} ||\mathbf{l}^T\mathbf{C}_i^*\mathbf{l}||^2 , \tag{3.23}$$

which is a multivariate fourth-order polynomial in l_1, l_2, l_3. We notice that the cost function admits the trivial solution $\mathbf{l} = \mathbf{0}$. The solution for $\mathbf{l}$ is given by

$$\hat{\mathbf{l}} = \arg\min_{\mathbf{l}} J\left(\mathbf{l}, \{\mathbf{C}_i\}_{i=0}^{M-1}\right). \tag{3.24}$$

Multiple Reflector Case

We shall conclude our analysis of the geometric constraint for reflector estimation by outlining the notation for the multiple reflector case. We do not, at this point, make any assumptions on the ordering of the TOA matrix $\boldsymbol{\tau}$, which is postponed until the next chapter. We extend the matrix decomposition of the ellipse given by (3.16) to the case where $\mathbf{C}_{i,k}$ is associated with the ith microphone ($i \in \{0, \cdots, M-1\}$) and the kth reflector ($k \in \{1, \cdots, K\}$) such that

$$\mathbf{C}_{i,k} = \mathbf{T}_i^{-T}\mathbf{R}_i^{-T}\mathbf{S}_{i,k}^{-T}\mathbf{C}_{\mathrm{I}}\mathbf{S}_{i,k}^{-1}\mathbf{R}_i^{-1}\mathbf{T}_i^{-1}, \tag{3.25}$$

where the translation, rotation and scaling matrices are given by

$$\mathbf{T}_i = \begin{bmatrix} 1 & 0 & \Delta x_i \\ 0 & 1 & \Delta y_i \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.26}$$

$$\mathbf{R}_i = \begin{bmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.27}$$

$$\mathbf{S}_{i,k} = \begin{bmatrix} Q_{i,k}^{\mathrm{smaj}} & 0 & 0 \\ 0 & Q_{i,k}^{\mathrm{smin}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.28}$$

Note that only the scaling of the semi-major and semi-minor axes of each ellipse changes such that

$$\begin{aligned} Q_{i,k}^{\mathrm{smaj}} &= \frac{\mathfrak{c} \cdot \tau_{i,k}}{2}; \\ Q_{i,k}^{\mathrm{smin}} &= \frac{\sqrt{(\mathfrak{c} \cdot \tau_{i,k})^2 - D_i^2}}{2}. \end{aligned}$$

Accordingly, the implicit equation of an ellipse with foci in $(x_{\mathrm{s}}, y_{\mathrm{s}})$ and (x_i, y_i) and with major diameter $r_{i,k} = \mathfrak{c}\tau_{i,k}$ is

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} + \sqrt{(x - x_{\mathrm{s}})^2 + (y - y_{\mathrm{s}})^2} = r_{i,k}\,. \tag{3.29}$$

By expanding (3.29) and comparing term by term with a conic with parameters $(a_{i,k}, b_{i,k}, c_{i,k}, d_{i,k}, e_{i,k}, f_{i,k})$ we obtain

$$\begin{aligned}
a_{i,k} &= r_{i,k}^2 - (x_\mathrm{s} - x_i)^2 \,, \\
b_{i,k} &= -2(x_\mathrm{s} - x_i)(y_\mathrm{s} - y_i) \,, \\
c_{i,k} &= r_{i,k}^2 - (y_\mathrm{s} - y_i)^2 \,, \\
d_{i,k} &= (x_i - x_\mathrm{s})(x_i^2 + y_i^2 - x_\mathrm{s}^2 - y_\mathrm{s}^2 - r_{i,k}^2) - x_\mathrm{s} r_{i,k}^2 \,, \\
e_{i,k} &= (y_i - y_\mathrm{s})(x_i^2 + y_i^2 - x_\mathrm{s}^2 - y_\mathrm{s}^2 - r_{i,k}^2) - y_\mathrm{s} r_{i,k}^2 \,, \\
f_{i,k} &= (x_\mathrm{s}^2 r_{i,k}^2 + y_\mathrm{s}^2 r_{i,k}^2) \frac{y_i^2 - y_\mathrm{s}^2 + x_i^2 - x_\mathrm{s}^2 - r_{i,k}^2}{4} \,.
\end{aligned}$$

Consequently, the cost function to estimate the line parameters of a particular reflector $k \in \{1, \cdots, K\}$ is given by

$$J_\mathrm{m}\left(\mathbf{l}, \{\mathbf{C}_{i,k}\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} \left\| \mathbf{l}^T \cdot \mathrm{adj}(\mathbf{C}_{i,k}) \cdot \mathbf{l} \right\|^2, \tag{3.30}$$

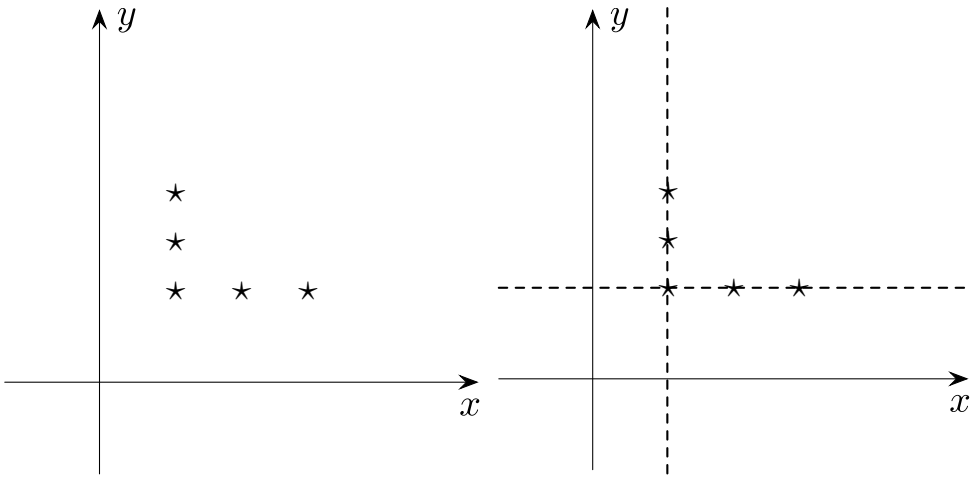
such that

$$\hat{\mathbf{l}}_k = \arg\min_{\mathbf{l}} J_\mathrm{m}\left(\mathbf{l}, \{\mathbf{C}_{i,k}\}_{i=0}^{M-1}\right). \tag{3.31}$$

In a noise-free scenario, and neglecting the effects of machine precision, the global minimum of (3.31) is also the true solution, so that all ellipses are perfectly aligned and yield a single solution that is the common tangent to all ellipses considered. However, due to errors in the TOA information, the ellipses are prone to mismatch and the solution from (3.31) is not guaranteed to correspond to the true reflector. In the next Section we present a method that robustly estimates the line parameters of a reflector.

3.4 Localization Using the Hough Transform

The Hough technique is particularly useful for computing a global description of a feature (where the number of solution classes need not be known a priori), given (possibly noisy) local measurements. The motivating idea behind the Hough technique for line detection is that each input measurement (i.e. coordinate point) indicates its contribution to a globally consistent solution (i.e. the physical line which gave rise to that image point).



(a) Coordinate points (b) Possible straight line fitting

Figure 3.6: Coordinate points and and possible straight line fitting.

As a simple example, consider the common problem of fitting a set of line segments to a set of discrete image points (i.e. pixel locations obtained from an edge detector on an image). Figure 3.6 shows a possible solution to this problem. Here the lack of *a priori* knowledge about the number of desired line segments renders this problem under-constrained and motivates the Hough transform approach.

The localization accuracy of piece-wise estimated reflectors can be improved by introducing spatial variation in the position of the sound source (or equivalently the microphones) and taking measurements at each step. Since the rigorous proof is beyond the scope of this work, an intuitive justification will be given in the following.

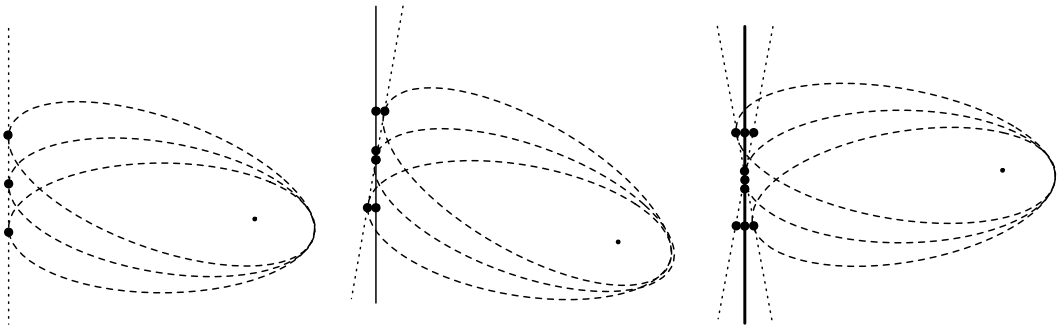
With slight abuse of notation we define a random variable $X \in \mathbb{R}^3$ that represents the three parameters of a particular line estimate from (3.31). By repeating the reflector localization for N source positions, we claim that X converges almost surely to the expected value, i.e. the true reflector parameters, by referring to the strong law of large numbers (LLN) [146]. In other words,

$$\bar{X}_N \overset{a.s.}{\to} \mu \quad \text{when } N \to \infty,$$

where $X_1, X_2, \cdots, X_N$ is an infinite sequence of i.i.d. integrable random variables with expected value $E\{X_1\} = E\{X_2\} = \cdots = \mu$. That is,

$$\Pr\left(\lim_{N\to\infty} \bar{X}_N = \mu\right) = 1.$$

The Hough transform, that acts on a sparse data point representation of localized



(a) $N = 1$ (b) $N = 2$ (c) $N = 3$

Figure 3.7: Increasing the measurements, or the *evidence*, one can observe intuitively that the relector-line estimates cluster around the true solution (note the increasing line weighting).

reflectors (more on this in the next Section), introduces an averaging effect (or smoothing) on the observed data, as seen in Figure 3.7. It will be shown using numerical simulations in Section 4.5.4 how this increases the robustness to (Gaussian) noise in the TOA information.

Quantifying the noise sources and distributions is beyond the scope of this work. Errors are introduced in the system in terms of (Gaussian) TOA measurement noise and sensor noise, amongst others. Recently, it was shown that the Hough transform is implicitly a Bayesian process [147]. This important fact allows the framework presented in this work to be connected with other Bayesian techniques in a well-grounded and principled way. In fact, shape detection can easily be regarded as a clustering problem: assigning features to generating distributions. This suggests that, for an unknown number of shapes, an infinite mixture model, which allows a prior distribution on the number of shapes, may be appropriate [147].

Additionally, given the connection of the Hough transform with a Bayesian process, and by adopting a hierarchical Bayesian network structure, in which the system learns the parameters and number of classes in an unsupervised way online from the observed data, one can motivate a process that offers the following two advantages:

1. The posterior distribution can be used to provide confidence information of the observed data (i.e. “error bars”).

2. No *a priori* assumptions on the number of classes means that more complex room geometries can potentially be taken into account and tied in with the existing inference framework.

3.4.1 Relationship Between Line Estimates and Ellipses

The input measurements (i.e. coordinate points) are created based on the geometric relationship between the set of ellipses and the line parameters of the most feasible reflector for that set. In other words, TOA information is "mapped" to points in the Hough space. Rather than establishing a direct parametrization of the TOAs we adhere to the geometric framework of this chapter. Given M microphones and N source positions, the aim is to define a set of candidate points to be used for the refinement of the first estimate, i.e. from the case $N = 1$. These points are defined as

$$\mathbf{p}_j \triangleq [x_j \; y_j]^T, \quad j = 0, \ldots, P, \tag{3.32}$$

where $M\,N - 1 \leq P \leq 2\,M\,N - 1$. The elements in $\mathbf{P} = [\mathbf{p}_0\,\mathbf{p}_1\,\cdots\,\mathbf{p}_P]$ are either points of intersection, points of tangency, or closest coordinate points of an ellipse and the initial reflector line estimate $\mathbf{l}$. Consequently, for every ellipse $\mathbf{C}$ and reflector line $\mathbf{l}$ the following hold:

- If $\mathbf{l}$ goes through $\mathbf{C}$ then we obtain two points of intersection.
- If $\mathbf{l}$ touches $\mathbf{C}$ at one point, or in other words if $\mathbf{l}$ is tangent to $\mathbf{C}$, then we obtain one point of tangency.
- If $\mathbf{l}$ does not go through $\mathbf{C}$ then we need to calculate the closest point on the line with respect to the conic.

3.4.2 Analytical Framework

In homogenous coordinates, any line $\mathbf{l}$ cutting through the ellipse $\mathbf{C}$ intersects the ellipse at the two points of intersection $\mathbf{p}_\alpha \triangleq [x_\alpha \; y_\alpha \; 1]^T$ and $\mathbf{p}_\beta \triangleq [x_\beta \; y_\beta \; 1]^T$. Furthermore, there exist two lines parallel to $\mathbf{l}$, i.e. with slope $m = -\frac{l_1}{l_2}$, that touch the ellipse at the tangential points $\mathbf{p}_{\bar{\alpha}} \triangleq [x_{\bar{\alpha}} \; y_{\bar{\alpha}} \; 1]^T$ and $\mathbf{p}_{\bar{\beta}} \triangleq [x_{\bar{\beta}} \; y_{\bar{\beta}} \; 1]^T$. Therefore the problem is

constrained to finding the points on the ellipse for which the tangents have slope m. This can be achieved by implicit differentiation of the ellipse, given by (3.11)

$$\frac{d}{dx}(\mathcal{C}) = 2\,a\,x + 2\,b\,y + 2\,b\,x\frac{dy}{dx} + 2\,c\,y\frac{dy}{dx} + 2\,d + 2\,e\frac{dy}{dx} = 0\,. \tag{3.33}$$

After setting $\frac{dy}{dx} = m$ the line that goes through both tangential points can be expressed as

$$\mathbf{l}_{\mathrm{T}} = [(a + b\,m)\,(b + c\,m)\,(d + e\,m)]^T. \tag{3.34}$$

For any line $\mathbf{l}$ it is possible to find the two points $\mathbf{p}_{T_\alpha}$ and $\mathbf{p}_{T_\beta}$ at which two lines are both parallel to $\mathbf{l}$, i.e. with slope m, and also tangential to the ellipse. Since we can construct the line $\mathbf{l}_{\mathrm{T}}$ that goes through both points $\mathbf{p}_{T_\alpha}$ and $\mathbf{p}_{T_\beta}$ from (3.34), what remains is to compute the points of intersection of $\mathbf{l}_{\mathrm{T}}$ and the ellipse. First, the methodology used for finding the general intersection points of a line and an ellipse is elaborated and it is then shown how the points $\mathbf{p}_{T_\alpha}$ and $\mathbf{p}_{T_\beta}$ can be computed.
Given a line $\mathbf{l}$ that goes through the ellipse $\mathbf{C}$, the points of intersection $\mathbf{p}_\alpha \triangleq [x_\alpha\ y_\alpha]^T$ and $\mathbf{p}_\beta \triangleq [x_\beta\ y_\beta]^T$ are given by

$$x_\alpha = \frac{l_2\sqrt{(A+B+C)}+D}{E}\,, \quad y_\alpha = -\frac{-l_3 + l_1\,x_\alpha}{l_2}\,; \tag{3.35}$$

$$x_\beta = -\frac{l_2\sqrt{(A+B+C)}-D}{E}\,, \quad y_\beta = -\frac{-l_3 + l_1\,x_\beta}{l_2}\,; \tag{3.36}$$

with

$$\begin{aligned}
A &= b\left(b\,l_3^2 - 2\,d\,l_2\,l_3 - 2\,e\,l_1\,l_3 + 2\,f\,l_1\,l_2\right)\,,\\
B &= d\left(d\,l_2^2 - 2\,e\,l_1\,l_2 + 2\,c\,l_1\,l_3\right)\,,\\
C &= e^2\,l_1^2 + 2\,a\,e\,l_2\,l_3 - c\,f\,l_1^2 - a\,f\,l_2^2 - a\,c\,l_3^2\,,\\
D &= b\,l_2\,l_3 - d\,l_2^2 - c\,l_1\,l_3 + e\,l_1\,l_2\,,\\
E &= c\,l_1^2 - 2\,b\,l_1\,l_2 + a\,l_2^2\,.
\end{aligned}$$

Instead of using $\mathbf{l}$ in (3.35) and (3.36) to find the general solutions $\mathbf{p}_\alpha$ and $\mathbf{p}_\beta$, we can replace $[l_1\ l_2\ l_3]^T$ with $[l_{\mathrm{T}_1}\ l_{\mathrm{T}_2}\ l_{\mathrm{T}_3}]^T$, as given by (3.34), such that

$$\mathbf{l} \triangleq [(a + b\,m)\,(b + c\,m)\,(d + e\,m)]^T, \tag{3.37}$$

in order to find $\mathbf{p}_{T_\alpha}$ and $\mathbf{p}_{T_\beta}$. Since any line $\mathbf{l}$ will have two parallel lines that are tangential to the ellipse, (3.35), (3.36) and (3.37) can be used to check whether $\mathbf{l}$ goes through the ellipse, is tangential to the ellipse or does not go through the ellipse.

Remark 1. *If* $\mathbf{l}$ *cuts through the ellipse, then there exists a line parallel to* $\mathbf{l}$ *that touches the ellipse at point* $\mathbf{p}_{\bar{\alpha}}$ *and is either to the left or right, above or below* $\mathbf{l}$*.*

Proof. Substitution into (3.10) yields $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} \neq 0$. If $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} > 0$ and $\mathbf{l}$ passes through the ellipse, then by definition $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} < 0$. Consequently if $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} < 0$ then $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} > 0$. □

Remark 2. *If* $\mathbf{l}$ *is tangential to the ellipse, then* $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} = 0 \vee \mathbf{l}^T\mathbf{p}_{\bar{\alpha}} \neq 0$*.*

Proof. If $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} = 0$, then $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} \neq 0$. Similarly if $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} = 0$, then $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} \neq 0$. □

Remark 3. *If* $\mathbf{l}$ *neither intersects or is tangential to the ellipse, then the two parallel lines touching the ellipse at points* $\mathbf{p}_{\bar{\alpha}}$ *and* $\mathbf{p}_{\bar{\beta}}$ *are either both below, above, left or right of* $\mathbf{l}$*.*

Proof. If $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} > 0$ then $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} > 0$. If $\mathbf{l}^T\mathbf{p}_{\bar{\alpha}} < 0$ then $\mathbf{l}^T\mathbf{p}_{\bar{\beta}} < 0$. □

Consequently, in order to determine the relationship between $\mathbf{l}$ and the ellipse, it is sufficient to compute

$$\Phi = \left|\operatorname{sgn}\left(\mathbf{l}^T\mathbf{p}_{\bar{\alpha}}\right) + \operatorname{sgn}\left(\mathbf{l}^T\mathbf{p}_{\bar{\beta}}\right)\right|, \tag{3.38}$$

where $\operatorname{sgn}(\,\cdot\,)$ is defined as

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0\,, \\ 0 & \text{if } x = 0\,, \\ 1 & \text{if } x > 0\,. \end{cases}$$

If $\Phi < 1$, then $\mathbf{l}$ goes through the ellipse. If $\Phi = 1$, then $\mathbf{l}$ is tangential to the ellipse. Finally, if $\Phi > 1$, then $\mathbf{l}$ does not intersect the ellipse.

3.4.3 Obtaining Candidate Points

New candidate points are appended to $\mathbf{P}$ in the following way. First (3.38) is used to classify the line into one of the three classes. In the first case it is sufficient to calculate the two points of intersection and add the resulting points to $\mathbf{P}$. In the second case the single point of intersection is calculated to obtain one point of tangency and the result stored in $\mathbf{P}$. In the final case the two tangential points $\mathbf{p}_{\bar{\alpha}}$ and $\mathbf{p}_{\bar{\beta}}$ are used since one of them will be the closest point on the line to the ellipse and the other the furthest. Since we are only interested in the closest point, it is sufficient to compute the distance

of points $\mathbf{p}_{\bar{\alpha}}$ and $\mathbf{p}_{\bar{\beta}}$ and the line, by projecting them both onto the line and selecting the shortest distance from

$$\min\left\{\frac{\left|l_1\,x_{\bar{\alpha}}+l_2\,y_{\bar{\alpha}}+l_3\right|}{\sqrt{l_1^2+l_2^2}},\frac{\left|l_1\,x_{\bar{\beta}}+l_2\,y_{\bar{\beta}}+l_3\right|}{\sqrt{l_1^2+l_2^2}}\right\}, \tag{3.39}$$

and adding the corresponding coordinate point to $\mathbf{P}$.

Discussion and Conclusion

A geometrical framework based on a study on projective geometry has been presented, outlining a solution to the reflector localization problem that is used in the remainder of this work. The problem can be summarized as follows: the reflective boundaries of an acoustic enclosure are located based upon TOAs estimated from acoustic measurements. The TOA information is used to form a set of elliptical constraints on the possible locations of the reflectors. The common tangent of these constraints corresponds to the reflector location that can be found by minimizing the objective function in (3.30) in a least-squares sense. Initially an acoustic scene containing a single planar reflector is considered. The cost function, whose minimum in an error-free scenario corresponds to the line parameters of the reflector, was then extended to the case of a more complex acoustic scene in which multiple reflectors are present. Finally, the adoption of the Hough technique, for increasing the robustness to Gaussian noise in the TOA information, was motivated.

4

Room Geometry Estimation in Two-Dimensions

Introduction

THE localization of reflectors, such as the walls enclosing a room, in a two-dimensional (2-D) plane is presented in detail in this chapter. Specifically, localization of multiple reflectors is achieved by estimation of the time-of-arrival (TOA) of reflected signals by analysis of acoustic impulse responses (AIRs). When multiple walls are present in the acoustic scene, an ambiguity problem arises, which can be addressed using the Hough transform.

The reflector localization method can be summarised as follows: i) AIRs are obtained by either supervised or unsupervised identification, with known or unknown input stimulus; ii) TOAs are estimated from the estimated AIRs; iii) TOA information, along with relative sensor positions, can be used to establish a geometric constraint; iv) the geometric constraint can be used to estimate possible reflector locations in a least-squares sense. An overview of the different methods outlined in this chapter, is depicted in Figure 4.1.

The remainder of this chapter is organized as follows: Section 4.1 outlines the system identification process. Specifically, the measurement and estimation of TOAs from unsynchronized AIRs, obtained through either supervised, or unsupervised identification, is discussed. Section 4.2 presents the common tangent estimation algorithm using a constrained LS estimator and an optimal closed-form estimator. Additionally,

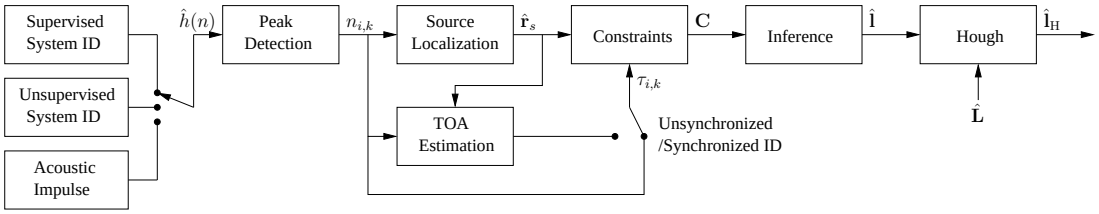


Figure 4.1: System diagram. An acoustic impulse response is generated, containing peaks corresponding to sound from reflected boundaries. Peak picking from either the true or estimated system yields TOAs if the measurements are synchronized, else TOAs are estimated with the aid of source localization. TOAs, combined with an estimate of the source and knowledge of the geometry of the receiver array, are used to parameterize a set of ellipses. Geometric inference is performed with the ellipses by finding lines of common tangency that correspond to the reflector locations.

the performance of the two estimators is compared theoretically and experimentally. Section 4.3 introduces the ambiguity problem in the TOA information matrix and proposes possible resorts. In particular, a representation based on the Hough transform is developed. Adopting such framework offers advantages in both addressing TOA ambiguity, and correction of reflector location data. The performance of the proposed inference algorithm is evaluated through simulations, and by measurements made in real conference rooms, in Section 4.5.

Relevant Publications

1. J. Filos, E. A. P. Habets and P. A. Naylor, "A Two-Step Approach To Blindly Infer Room Geometries," *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, August 30 - September 2, 2010.

2. J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti and P. A. Naylor, "Robust Inference of Room Geometry from Acoustic Impulse Responses," *Proc. European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, August 29 - September 2, 2011.

3. A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor and S. Tubaro, "Exact Localization of Acoustic Reflectors from Quadratic Constraints," *Proc. Workshop on App. of Signal Processing to Audio and Acoust. (WASPAA)*, New Paltz, New York, USA, October 16 - 19, 2011.

4. F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor and S. Tubaro, "Inference of Room Geometry from Acoustic Impulse Responses," *IEEE Trans. Audio, Speech, Lang. Process.*, 2012.

4.1 Measurement and Estimation of TOAs

In this Section we illustrate the steps that, starting from the AIRs, lead to the estimation of TOAs. In particular, we consider both the cases of synchronized and unsynchronized AIRs. In the former case the TOAs are directly extracted from the impulse response. In the latter situation, instead, TOAs can be estimated once the source has been localized. In order to make the peak detection algorithm robust against non-ideal acquisition and emission systems, we propose a template matching procedure to improve the relevant temporal characteristics of the received source signal.

4.1.1 Estimation of TOAs from Unsynchronized AIRs

In order to estimate TOAs from unsynchronized AIRs, the time-difference-of-arrivals (TDOAs) of the direct-paths are used to localize the acoustic source and consequently estimate the propagation time of the direct sound from the source to a reference microphone. The propagation times of all the other arrivals can then be inferred. As outlined in Section 2.3.2, and reproduced here for convenience, a least-squares estimator can be used to define the error function

$$\mathbf{e}(\mathbf{r}_\mathrm{s}) = \mathbf{A}\theta - \mathbf{b},$$

where

$$\mathbf{A} \triangleq \left[\mathbf{S} \mid \hat{\mathbf{d}}\right], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{M-1} & y_{M-1} \end{bmatrix},$$

$$\theta \triangleq \begin{bmatrix} x_\mathrm{s} \\ y_\mathrm{s} \\ R_\mathrm{s} \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{1,0}^2 \\ R_2^2 - \hat{d}_{2,0}^2 \\ \vdots \\ R_{M-1}^2 - \hat{d}_{M-1,0}^2 \end{bmatrix}.$$

The corresponding LS criterion is then given by

$$J = \mathbf{e}^T\mathbf{e} = [\mathbf{A}\theta - \mathbf{b}]^T [\mathbf{A}\theta - \mathbf{b}] .$$

The solution for θ is given by [80]

$$\hat{\theta}_1 = \mathbf{A}^\dagger \mathbf{b},$$

We can now use the estimate of the distance from the ith microphone to the source in order to estimate the TOAs of the direct-path for each of these microphones

$$\hat{\tau}_{i,0} = \frac{\hat{D}_i}{\mathfrak{c}}, \quad i = 0, \ldots, M-1. \tag{4.1}$$

The TOAs of the reflective paths can be obtained from the above equation since both $\hat{\tau}_{i,0}$ and the TDOAs between the direct-paths and the reflective paths are known from inspection of $h_i(t)$, even if the source and microphone signals are not synchronized.

4.1.2 Blind Identification of AIRs

AIRs measured in real acoustic environments present a challenging problem as the source impulse-like emission $h^{\mathrm{s}}(n)$ is convolved with the AIR $h_i(n)$. One of the challenges is accounting for fractional delays resulting from path lengths that are not multiples of the distance propagated by sound in one sample period. Detection of impulsive events can be achieved to within one sample by considering local centres of energy with algorithms such as the sliding group delay function [148] and the *findpeaks* function [149].

4.1.2.1 Template Matching Procedure

Assuming supervised identification with which estimation error can be ignored, the measured AIR is

$$\hat{h}_i(n) = h^{\mathrm{s}}(n) * h_i(n). \tag{4.2}$$

An example impulse response for a measured system is seen in Figure 4.2 (a), showing respectively the direct-path and three first-order reflections for a single channel. The centres of each event are marked by '$\circ$', each of which are surrounded by nearby ripples caused by $h^{\mathrm{s}}(n)$. The ripples cause uncertainty in determining the exact time corresponding to the peak and therefore a matched filter was proposed in [150] to alleviate this problem.

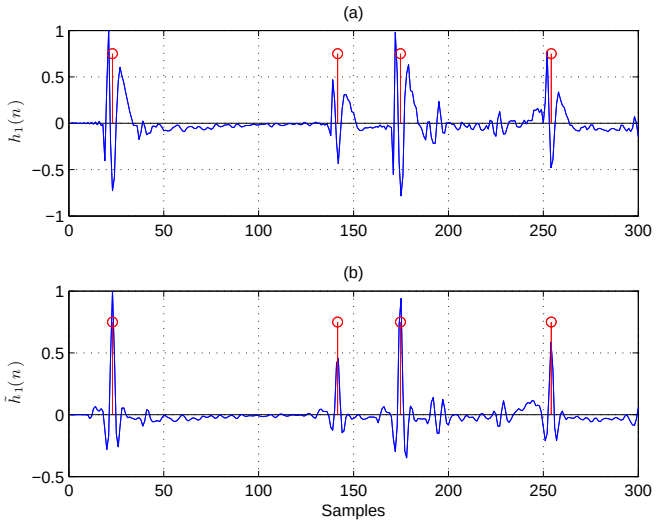


Figure 4.2: Direct-path and three first-order reflections for (a) measured impulse response, (b) modified impulse response according to (4.4). Red '○' mark the estimated peak locations.

The length of $h^{\mathrm{s}}(n)$ is usually sufficiently short that it has decayed before the arrival of the first-order reflections [151], as in the example. Therefore, $h^{\mathrm{s}}(n)$ can be observed from the first few nonzero taps in $\hat{h}_i(n)$. Let n_i^{DP} be the propagation time of the direct-path signal from the source to microphone i and N^{s} be the approximate length of the loudspeaker impulse response. An impulsive source emission, such as a hand-clap, can be estimated by $\hat{h}_i^{\mathrm{s}}(n) = \hat{h}_i(n + n_i^{\mathrm{DP}})\, w_i(n)$, where

$$w_i(n) = \begin{cases} 1 & \text{if } 0 \leq n < N^{\mathrm{s}}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

The filter $\hat{h}^{\mathrm{s}}(n)$ is equalized through the sliding correlation or matched filter [150],

$$\tilde{h}_i(n) = \sum_{j=0}^{N^{\mathrm{s}}-1} \hat{h}_i^{\mathrm{s}}(j)\, h_i(n+j), \tag{4.4}$$

that equalizes $\hat{h}_i^{\mathrm{s}}$ to a single peak as demonstrated in Fig. 4.2 for a measured AIR. In (b) the mean group delay of $\hat{h}_i^{\mathrm{s}}$ has been compensated. The detected peaks are denoted by $n_{i,k}$ where i and k are the microphone and reflector index respectively.

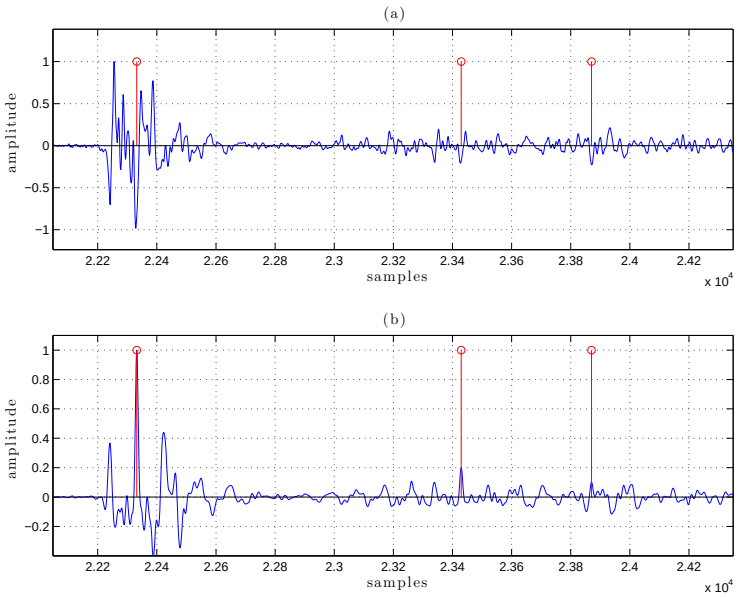


Figure 4.3: Capturing a hand clap: (a) impulse-like input stimulus, (b) filtered output. The red '○' mark the true location.

Template Matching Procedure for Quasi-AIRs

Many existing geometric approaches rely upon the availability of acoustic impulse responses that may be impractical to obtain in a domestic setting. Supervised methods such as MLS and TSP require 1) a loudspeaker system emitting the source sequence; 2) high output signal level to achieve adequate dynamic range.

For this reason we propose the use of uncontrolled impulsive acoustic sources, such as a hand clap or the snapping of fingers, to obtain quasi-AIRs that approximate (up to a certain number of taps) a true AIR. An example recording for a measured system is seen in Figure 4.3 (a), showing respectively the direct-path and two first-order reflections for a single channel. The estimated time instant of each event is marked by a red '○', each of which is surrounded by nearby ripples from $h^{\mathrm{s}}(n)$. The relatively large temporal support of such acoustic stimuli presents an additional challenge as it introduces uncertainty as to the exact temporal location of the peaks in the received signal that correspond to both the direct-path and reflected TOAs. A preprocessor, based on a matched filter to reduce the temporal support of the peaks in the quasi-AIRs, is applied as a first step. The filter $\hat{h}^{\mathrm{s}}(n)$ could be equalized by finding a filter

$g^{\mathrm{s}}(n)$, such that $\hat{h}^{\mathrm{s}}(n) * g^{\mathrm{s}}(n) \simeq \delta(n)$ in a least-squares sense. Although this is an optimal solution [150] it has been found unreliable in most practical situations. The suboptimal but more practical (in practice) approach is again given by the matched filter approach in (4.4), i.e. $\tilde{h}_i(n) = \sum_{j=0}^{N^{\mathrm{s}}-1} \hat{h}_i^{\mathrm{s}}(j) h_i(n+j)$.

Figure 4.3 demonstrates this for a measured AIR showing the direct-path and two first-order peaks in (a) $\hat{h}_1(n)$; and (b) $\tilde{h}_1(n)$, where the mean group delay of $\hat{h}_i^{\mathrm{s}}$ has been compensated.

4.2 Common Tangent Estimation

In this section we present three possible approaches to estimate the line parameters of the common reflector between groups of microphones. First, the constrained least-squares solution is presented, that is not guaranteed to give the optimal solution. As a next step, a closed-form estimator is presented that solves the problem in an analytic way. Finally, an equivalent exact solution is presented that reformulates the optimal estimator.

4.2.1 Constrained Least-Squares Solution

As noted in the previous chapter, the cost function (3.23), i.e.

$$J\left(\mathbf{l}, \{\mathbf{C}_i\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} ||\mathbf{l}^T \mathbf{C}_i^* \mathbf{l}||^2 \,,$$

is a multivariate fourth-order polynomial in l_1, l_2, l_3 and admits the trivial solution $\mathbf{l} = \mathbf{0}$. This cost function can be solved using an iterative least-squares estimator such that

$$\hat{\mathbf{l}} = \arg\min_{\mathbf{l}} J\left(\mathbf{l}, \{\mathbf{C}_i\}_{i=0}^{M-1}\right).$$

The objective function is non-convex since the ellipse matrix is not positive definite or semidefinite. Consequently, when employing an optimization algorithm to find $\mathbf{l}$, it is possible to get trapped in a local minimum rather than finding the global minimum. As proposed in [113], this problem can be alleviated by imposing that l_1 and l_2 lie on a circle of radius 1:

$$l_1 = \cos(\alpha),\ l_2 = \sin(\alpha). \tag{4.5}$$

We can find $\hat{\mathbf{l}}_\alpha = [\cos(\alpha), \sin(\alpha), l_3]^T$ by minimizing J in (3.23) using $\mathbf{l}_\alpha$ rather than $\mathbf{l}$

$$J\left(\mathbf{l}_\alpha, \{\mathbf{C}_i\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} ||\mathbf{l}_\alpha^T \mathbf{C}_i^* \mathbf{l}_\alpha||^2 \,, \tag{4.6}$$

such that

$$\hat{\mathbf{l}}_\alpha = \arg\min_{\mathbf{l}_\alpha} J\left(\mathbf{l}_\alpha, \{\mathbf{C}_i\}_{i=0}^{M-1}\right). \tag{4.7}$$

4.2.2 Closed-Form Solution

In order to find the global minimum we resort to an analytical minimization technique [115] by slicing the homogeneous coordinates space (l_1, l_2, l_3) with the three planes $l_1 = 1$, $l_2 = 1$ and $l_3 = 1$. On these planes the cost function $J(\mathbf{l})$ is not homogeneous and the set of local minima can be found in an analytical way. By merging the minima found on the three planes we obtain the global solution.

We denote the coefficients of the adjoint conic associated to the ith ellipse with the matrix

$$\mathbf{C}_i^* = \begin{bmatrix} \alpha_i & \beta_i/2 & \delta_i/2 \\ \beta_i/2 & \gamma_i & \varepsilon_i/2 \\ \delta_i/2 & \varepsilon_i/2 & \zeta_i \end{bmatrix} .$$

Using this notation the cost function can be expanded as

$$\begin{aligned} J(\mathbf{l}) \;=\; & \sum_{i=0}^{M-1} [\alpha_i^2 l_1^4 + \gamma_i^2 l_2^4 + \zeta_i^4 l_3^4 + 2\alpha_i\beta_i l_1^3 l_2 + 2\alpha_i\delta_i l_1^3 l_3 + 2\beta_i\gamma_i l_1 l_2^3 + \\ & + 2\gamma_i\varepsilon_i l_2^3 l_3 + 2\delta_i\zeta_i l_1 l_3^3 + 2\varepsilon_i\zeta_i l_2 l_3^3 + (2\alpha_i\gamma_i + \beta_i^2) l_1^2 l_2^2 + \\ & + (2\alpha_i\zeta_i + \delta_i^2) l_1^2 l_3^2 + (2\gamma_i\zeta_i + \varepsilon_i^2) l_2^2 l_3^2 + 2(\alpha_i\varepsilon_i + \beta_i\delta_i) l_1^2 l_2 l_3 + \\ & + 2(\beta_n\varepsilon_n + \gamma_n\delta_n) l_1 l_2^2 l_3 + 2(\beta_n\zeta_n + \delta_n\varepsilon_n) l_1 l_2 l_3^2] \,. \end{aligned} \tag{4.8}$$

Slicing $J(\mathbf{l})$ with the planes $l_1 = 1$, $l_2 = 1$ and $l_3 = 1$ means computing $J(\mathbf{l})|_{l_1=1}$, $J(\mathbf{l})|_{l_2=1}$ and $J(\mathbf{l})|_{l_3=1}$, respectively.

We proceed by finding the zeros of the gradient of $J(\mathbf{l})|_{l_1=1}$, $J(\mathbf{l})|_{l_2=1}$ and $J(\mathbf{l})|_{l_3=1}$, so that we obtain the sets

$$L_1 = \left\{ \mathbf{l} : \left.\frac{\partial J(\mathbf{l})}{\partial l_2}\right|_{l_1=1} = 0 \;\wedge\; \left.\frac{\partial J(\mathbf{l})}{\partial l_3}\right|_{l_1=1} = 0 \right\} \,, \tag{4.9}$$

$$L_2 = \left\{ \mathbf{l} : \left. \frac{\partial J(\mathbf{l})}{\partial l_1} \right|_{l_2=1} = 0 \ \wedge \ \left. \frac{\partial J(\mathbf{l})}{\partial l_3} \right|_{l_2=1} = 0 \right\} , \tag{4.10}$$

and

$$L_3 = \left\{ \mathbf{l} : \left. \frac{\partial J(\mathbf{l})}{\partial l_1} \right|_{l_3=1} = 0 \ \wedge \ \left. \frac{\partial J(\mathbf{l})}{\partial l_2} \right|_{l_3=1} = 0 \right\} . \tag{4.11}$$

Notice that the partial derivatives of the slices $J(\mathbf{l})_{l_1=1}$, $J(\mathbf{l})_{l_2=1}$ and $J(\mathbf{l})_{l_3=1}$ are polynomials of order 3, and therefore L_1, L_2 and L_3 contain 9 solutions each. Some of them are in the complex domain and do not admit a solution. We denote with $\bar{L}_1$, $\bar{L}_2$ and $\bar{L}_3$ the subsets of purely real solutions of L_1, L_2 and L_3, respectively. We then define

$$\bar{L} = \bar{L}_1 \cup \bar{L}_1 \cup \bar{L}_3 = \{ \mathbf{l}_1 \dots \mathbf{l}_{K^{\text{CS}}} \} , \tag{4.12}$$

which contains $K^{\text{CS}} \leq 27$ candidate solutions. The global minimum of $J(\mathbf{l})$ is selected as

$$\hat{\mathbf{l}} = \arg \min_{\mathbf{l}_m} J(\mathbf{l}_m) \ , \ \mathbf{l}_m \in \bar{L} \ . \tag{4.13}$$

Notice that the trivial solution $\mathbf{l} = 0$ is inherently avoided by cutting the line parameter space with the planes $l_1 = 1$, $l_2 = 1$ and $l_3 = 1$. Note also that $J(\mathbf{l})|_{l_1=1}$, $J(\mathbf{l})|_{l_2=1}$ and $J(\mathbf{l})|_{l_3=1}$ are no longer homogeneous.

Finally, Figure 4.4 shows an example of slices $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$ (right-hand side), for the configuration of microphones and sources on the left-hand side. The correct line parameters for the configuration under analysis are $l_1 = -0.14, l_2 = -0.14, l_3 = 1$. We also notice that the minimum of $J(\mathbf{l})|_{l_3=1}$ has an asymmetric shape, and in particular is sharper along the radius of the circumference centered in $l_1 = 0, l_2 = 0$. As a consequence, the distance of the reflector will be identified better than its orientation. This fact depends on the mutual configuration of the source and the microphones and cannot be attributed to the proposed methodology.

4.2.3 Exact Solution

The geometric constraint used in the previous section is modified to derive an exact minimization of a constrained least-squares cost function. In particular, with reference to Figure 4.5, the image source $\mathbf{r}'_{\text{s}}$ is obtained by mirroring $\mathbf{r}_{\text{s}}$ over $\mathbf{l}$.

As shown in [113, 114], the TOA measures corresponding to the reflective paths can be converted into quadratic constraints (in the homogeneous space) describing an ellipse. More specifically, as shown in Figure 4.6, the ellipse has foci in $\mathbf{r}_i$ and $\mathbf{r}_{\text{s}}$, and

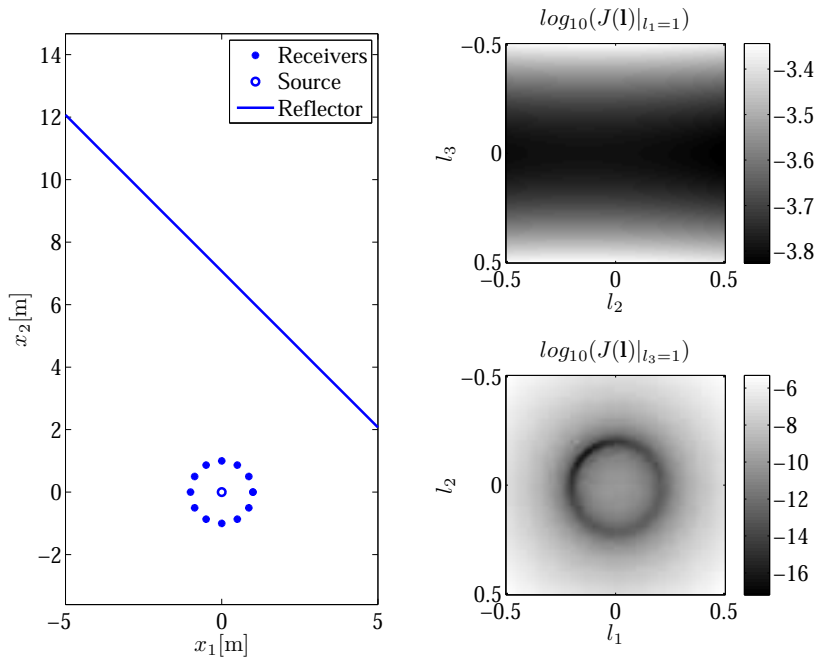


Figure 4.4: Example of cost functions $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$ for a specific configuration of microphones and sources [3].

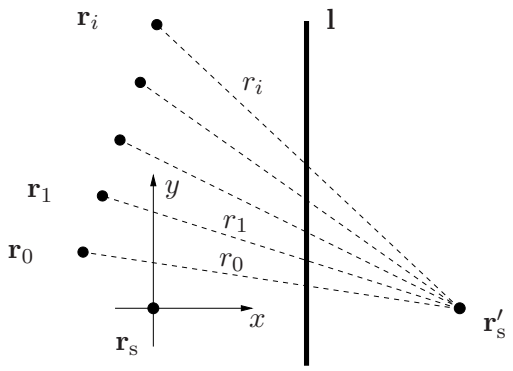


Figure 4.5: An acoustic source located at $\mathbf{r}_\mathrm{s}$ is reflected over the line $\mathbf{l}$ to its image position $\boldsymbol{s}'$. The microphones at $\mathbf{r}_0, \dots \mathbf{r}_{M-1}$ estimates their distances $r_0, \dots r_{M-1}$ from the image source in $\mathbf{r}'_\mathrm{s}$.

its major axis is r_i. This ellipse is tangential to the reflector line $\mathbf{l}$ at the reflection point $\mathbf{r}_{\mathrm{p},i}$. We modify the constraint on the implicit equation of the ellipse from the previous section such that

$$\sqrt{(x-x_i)^2+(y-y_i)^2} = r_i - \sqrt{x^2+y^2}\,. \tag{4.14}$$

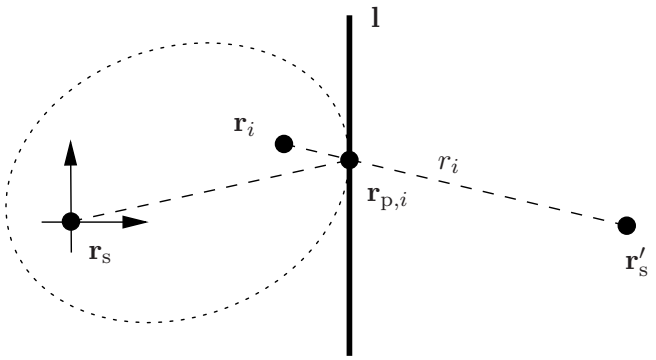


Figure 4.6: The length of the reflected path r_i from the image source in $\mathbf{r}'_\mathrm{s}$ to the microphone at $\mathbf{r}_i$ constrains the reflector line $\mathbf{l}$ to be tangent to an ellipse whose major axis is r_i and whose foci are $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$. $\mathbf{r}_{\mathrm{p},i}$ is the reflection point on $\mathbf{l}$.

After taking the square power of both sides of (4.14) we derive

$$\sqrt{x^2+y^2} - 2xx_i - 2yy_i = r_i^2 - x_i^2 - y_i^2 \ . \tag{4.15}$$

Squaring again both the sides of (4.15), we compare the implicit form of the ellipse, described by the homogeneous parameter vector $[a_i\ b_i\ c_i\ d_i\ e_i\ f_i]^T$, and given by

$$a_i x^2 + b_i x + c_i y^2 + d_i x + e_i y + f_i = 0 \ , \tag{4.16}$$

term-by-term to obtain

$$\begin{array}{rclcrcl} a_i &=& -4(r_i^2 - x_i^2) \ , & & d_i &=& 4[x_i r_i^2 - x_i(x_i^2+y_i^2)] \ , \\ b_i &=& 8x_i y_i \ , & & e_i &=& 4[y_i r_i^2 - y_i(x_i^2+y_i^2)] \ , \\ c_i &=& -4(r_i^2 - y_i^2) \ , & & f_i &=& r_i^4 - 2r_i^2(x_i^2+y_i^2) + (x_i^2+y_i^2)^2 \ . \end{array}$$

The implicit equation (4.16) can be expressed in matrix form as

$$\mathbf{x}^T \mathbf{C}_i \mathbf{x} = 0 \ , \ \mathbf{C}_i = \begin{bmatrix} a_i & b_i/2 & d_i/2 \\ b_i/2 & c_i & e_i/2 \\ d_i/2 & e_i/2 & f_i \end{bmatrix} , \tag{4.17}$$

where $\mathbf{x} = [\lambda x\ \lambda y\ \lambda]^T$ is the homogeneous representation of a point lying on the ellipse and $\mathbf{C}_i$ is the *point-conic* matrix. The dual form of the conic expresses the conic as line $\mathbf{l}$ tangential to it, i.e. $\mathbf{l}^T \mathbf{C}_i^* \mathbf{l} = 0$, where $\mathbf{l} = [l_1\ l_2\ l_3]^T$ is the homogeneous representation of a line tangent to the ellipse; and $\mathbf{C}_i^* = \det(\mathbf{C}_i)\mathbf{C}_i^{-1}$ represents the *line-conic* matrix.

In the following, we reformulate the cost function of Section 4.2.1. As noted in [152], such problems are referred to as Generalized Trust Region Subproblems (GTRS), whose

exact solution can be derived efficiently.

We first analyze the structure of the dual-conic, whose matrix

$$\mathbf{C}_i^* = \begin{bmatrix} \alpha_i & \beta_i/2 & \delta_i/2 \\ \beta_i/2 & \gamma_i & \varepsilon_i/2 \\ \delta_i/2 & \varepsilon_i/2 & \zeta_i \end{bmatrix},$$

is symmetric, and its parameters can be written as

$$\begin{array}{rclcrcl} \alpha_i^* & = & 4r_i^2(r_i^2 - x_i^2 - y_i^2)^2 \,, & & \delta_i^* & = & 16r_i^2 x_i (r_i^2 - x_i^2 - y_i^2) \,, \\ \beta_i^* & = & 0 \,, & & \varepsilon_i^* & = & 16r_i^2 y_i (r_i^2 - x_i^2 - y_i^2) \,, \\ \gamma_i^* & = & \alpha_i^* \,, & & \zeta_i^* & = & 16r_i^2 (r_i^2 - x_i^2 - y_i^2) \,. \end{array} \tag{4.18}$$

By substituting eq. (4.18) into the cost function (4.6), after some manipulation we obtain

$$J = \sum_{i=0}^{M-1} \left[\alpha_i^*(l_1^2 + l_2^2) + \delta_i^* l_1 l_3 + \varepsilon_i^* l_2 l_3 + \zeta_i^* l_3^2\right]^2 \,. \tag{4.19}$$

In order to find a unique minimum for J, we focus on the subspace defined by $\mathbf{l}' = [l_1 \; l_2 \; 1]^T$, and look for minima of the cost function lying on $l_3 = 1$. This leads to

$$\hat{\mathbf{l}}' = \arg\min_{\mathbf{l}'} \sum_{i=0}^{M-1} \left[\alpha_i^*(l_1^2 + l_2^2) + \delta_i^* l_1 + \varepsilon_i^* l_2 + \zeta_i^*\right]^2 \,. \tag{4.20}$$

Notice that the condition $l_3 = 1$ rules out the potential reflectors passing through the origin. As the origin is the location of the source, this does not constitute a serious limitation. The simple substitution $w = l_1^2 + l_2^2$ allows us to rewrite the vector of the unknowns as $\mathbf{w} = [w \; l_1^2 \; l_2^2]^T$, therefore the optimization problem can be written as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w} - \mathbf{b}\|^2 : \mathbf{w}^T\mathbf{D}\mathbf{w} + 2\mathbf{f}^T\mathbf{w} = 0 \right\} \,, \tag{4.21}$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha_0^* & \delta_0^* & \varepsilon_0^* \\ \vdots & \vdots & \vdots \\ \alpha_{M-1}^* & \delta_{M-1}^* & \varepsilon_{M-1}^* \end{bmatrix} \,, \; \mathbf{b} = \begin{bmatrix} -\zeta_0^* \\ \vdots \\ -\zeta_{M-1}^* \end{bmatrix},$$

and

$$\mathbf{D} = \mathrm{diag}(0, 1, 1) \,, \; \mathbf{f} = \begin{bmatrix} -0.5 & 0 & 0 \end{bmatrix}^T \,.$$

Assuming that $\mathbf{A}$ has full column rank, the problem can be solved efficiently, and the exact solution is readily found using the approach described in [152]. In particular, the minimum is found as

$$\hat{\mathbf{w}}(\lambda) = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D})^{-1}(\mathbf{A}^T\mathbf{b} - \lambda\mathbf{f}) \,,$$

where λ is the unique solution of $\hat{\mathbf{w}}(\lambda)^T\mathbf{D}\hat{\mathbf{w}}(\lambda)+2\mathbf{f}^T\hat{\mathbf{w}}(\lambda)=0$ on the interval for which $\mathbf{A}^T\mathbf{A}+\lambda\mathbf{D}$ is positive definite [152]. From the solution $\hat{\mathbf{w}}=[\hat{w}\ \hat{l}_1\ \hat{l}_2]^T$, the estimated reflector line is finally given by $\hat{\mathbf{l}}'=[\hat{l}_1\ \hat{l}_2\ 1]^T$.

4.2.4 Error Propagation Analysis

In this Section we propose a method for predicting the impact of the error on TOAs on the localization of reflectors using a formulation based on Catastrophe Theory [153]. Let $\mathbf{l}_0$ be the true reflector and $\mathbf{r}_0$ the true propagation distance of the reflective paths. In a real scenario the measurement of $\mathbf{r}_0$ is affected by an error expressed as $\delta\mathbf{r}$, such that noisy measurements are denoted by $\overline{\mathbf{r}}=\mathbf{r}_0+\delta\mathbf{r}$. Subsequently, the minimum of J becomes $\overline{\mathbf{l}}=\mathbf{l}_0+\delta\mathbf{l}$. Assuming the error $\delta\mathbf{r}$ to be sufficiently small, we want to find a relationship between $\delta\mathbf{r}$ and $\delta\mathbf{l}$. We do so by computing the second-order Taylor expansion of J centered about $(\mathbf{l}_0;\mathbf{r}_0)$. The term $(\nabla_{\mathbf{l}}J)^T|_{\mathbf{l}_0,\mathbf{r}_0}$ is zero, as the function with the true TOAs $\mathbf{r}_0$ has a minimum in $\mathbf{l}_0$. We can thus take the first-order derivative of the Taylor expansion and set it to zero to obtain

$$\mathbf{H}_{\mathbf{l},\mathbf{l}}(J)|_{\mathbf{l}_0,\mathbf{r}_0}\delta\mathbf{l}+\mathbf{H}_{\mathbf{l},\mathbf{r}}(J)|_{\mathbf{l}_0,\mathbf{r}_0}\delta\mathbf{r}=0\ , \tag{4.22}$$

where

$$\mathbf{H}_{\mathbf{l},\mathbf{l}}(J)=\begin{bmatrix}J_{l_1l_1} & J_{l_1l_2} & J_{l_1l_3}\\ J_{l_2l_1} & J_{l_2l_2} & J_{l_2l_3}\\ J_{l_3l_1} & J_{l_3l_2} & J_{l_3l_3}\end{bmatrix},\ \mathbf{H}_{\mathbf{l},\mathbf{r}}(J)=\begin{bmatrix}J_{l_1r_1} & \dots & J_{l_1r_N}\\ J_{l_2r_1} & \dots & J_{l_2r_N}\\ J_{l_3r_1} & \dots & J_{l_3r_N}\end{bmatrix},$$

and

$$J_{l_il_j}=\frac{\partial^2 J}{\partial l_i\partial l_j},\ \ J_{l_ir_j}=\frac{\partial^2 J}{\partial l_i\partial r_j}\ .$$

From (4.22) we finally obtain

$$\delta\mathbf{l}=\mathbf{G}\delta\mathbf{r}\ , \tag{4.23}$$

where $\mathbf{G}=-\mathbf{H}_{\mathbf{l},\mathbf{l}}(J)|^{-1}_{\mathbf{l}_0,\mathbf{r}_0}\cdot\mathbf{H}_{\mathbf{l},\mathbf{r}}(J)|_{\mathbf{l}_0,\mathbf{r}_0}$. In a real scenario we cannot assume $\delta\mathbf{r}$ to be known. However, some statistical information could be available in advance or could be estimated from the data. It is therefore important to find a relation between statistical descriptors of the noise $\delta\mathbf{r}$ and of $\delta\mathbf{l}$. The relationship between the covariance matrix $\mathbf{M}_{\mathbf{l}}$ of the estimation, and the covariance matrix $\mathbf{M}_{\mathbf{r}}$ of $\delta\mathbf{r}$ is

$$\mathbf{M}_{\mathbf{l}}=\mathbf{G}\mathbf{M}_{\mathbf{r}}\mathbf{G}^T, \tag{4.24}$$

where

$$\mathbf{M_l} = \begin{bmatrix} \sigma_{l_1}^2 & \sigma_{l_1}\sigma_{l_2} & \sigma_{l_1}\sigma_{l_3} \\ \sigma_{l_1}\sigma_{l_2} & \sigma_{l_2}^2 & \sigma_{l_2}\sigma_{l_3} \\ \sigma_{l_1}\sigma_{l_3} & \sigma_{l_2}\sigma_{l_3} & \sigma_{l_3}^2 \end{bmatrix}, \quad \mathbf{M_r} = \begin{bmatrix} \sigma_{r_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{r_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{r_N}^2 \end{bmatrix},$$

under the assumption of statistical independence of the measurement errors.

4.2.5 Comparison Between Exact and Iterative Methods

In order to compare the exact and iterative methods, numerical simulation were conducted with reference to the setup in Figure 4.7. The microphone array was made of

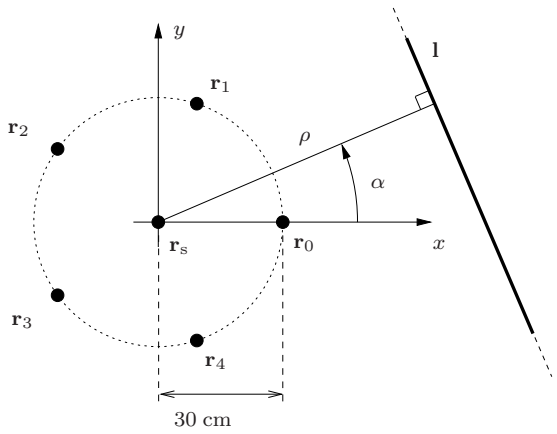
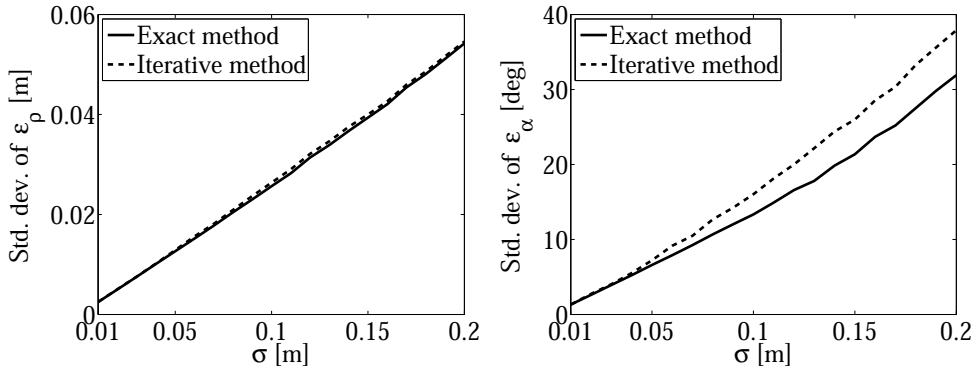


Figure 4.7: Simulation setup: the acoustic source is located in $\mathbf{r}_s$, corresponding to the centre of a 5-element circular microphone array. The line reflector $\mathbf{l}$ is described by its distance ρ and angle α from the origin.

5 sensors uniformly spaced on a circle of radius 30 cm centered in the origin of the reference frame (corresponding to the acoustic source). TOAs between microphones and source were calculated. The simulations were performed on a set of 9000 test reflector lines $\mathbf{l} = [\cos\alpha,\ \sin\alpha,\ -\rho]^T$ defined by their distance ρ and angle α with respect to the origin, as shown in Figure 4.7. The test reflectors were defined by distances in the range [1 m $\sim$ 4 m] and angles in the range [0 $\sim 2\pi$].

Using the above setup the performance of the exact and iterative methods for minimizing the cost function of Section 4.2.3 are compared. The iterative method considered for the comparison is enumerated in Section 4.2.1 [113, 114]. For each reflector position

errors were introduced into the distance measurements using 1000 realizations of independent identically distributed zero-mean Gaussian noise with standard deviation σ. The performance was evaluated by considering the distance error $\epsilon_\rho = |\rho - \hat{\rho}|$ and the angular error $\epsilon_\alpha = |\alpha - \hat{\alpha}|$ of the estimated reflector represented by the pair $(\hat{\rho}, \hat{\alpha})$ with respect to the true reflector position (ρ, α). Figs. 4.8-(a) and 4.8-(b) show the standard deviation of the distance error and of the angular error as a function of σ, respectively, averaged over all the tested locations and repetitions. As far as the distance error is



(a) Averaged std. dev. of ϵ_ρ (b) Averaged std. dev. of ϵ_α

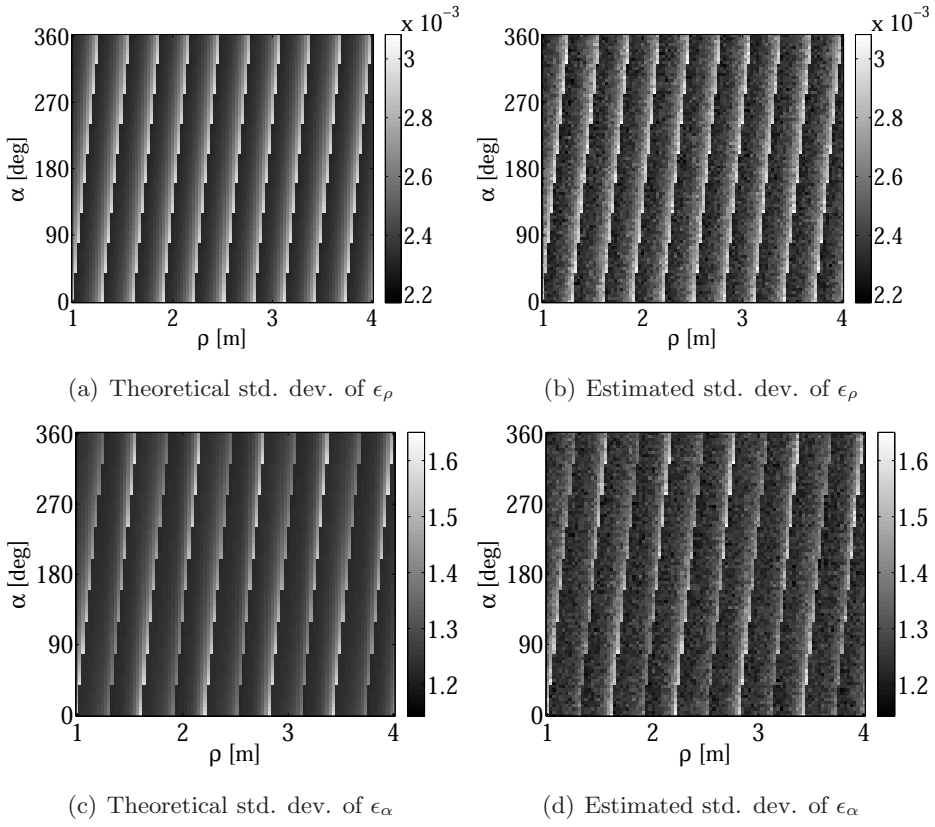
Figure 4.8: Comparison between the iterative and the exact solutions.

concerned, the iterative and the exact solutions exhibit almost identical errors, which were proportional to the standard deviation σ of the measurement error. As for the angular error, for values of σ below 0.05 m, the two approaches had virtually equivalent results, but for higher values of σ, the iterative method was affected by larger errors. This was due to the presence of multiple local minima in the cost function. For large measurement errors, the risk of encountering local minima increases as the cost function becomes less smooth. Although this phenomenon occurs for some error conditions, its impact on the standard deviation of the angular error is quite noticeable. The exact solution is therefore preferable over the iterative one, especially for large measurement errors.

4.2.6 Theoretical Error Analysis

We now validate the method for the error propagation analysis proposed in Section 4.2.4. In this case the standard deviation of the measurement noise is kept to $\sigma = 0.01$ m. The

standard deviation of the error predicted with the analytic method is compared with the results of the simulations conducted on the same testing reflector positions. The results shown in Figure 4.9 show the distance error for theoretical (a) and simulated (b) analysis, respectively. Similarly, Figure 4.9 shows the theoretical (c) and simulated (d) results relative to the angular error. The results of the simulations accurately match the



(a) Theoretical std. dev. of ϵ_ρ

(b) Estimated std. dev. of ϵ_ρ

(c) Theoretical std. dev. of ϵ_α

(d) Estimated std. dev. of ϵ_α

Figure 4.9: Comparison between the theoretical standard deviation of the error (predicted with the error propagation analysis) and simulation results.

theoretical ones: they present the same mean error of the expected values (2.5 mm for the distance and 1.3° for the angle). The patterns of local maxima (i.e. diagonal white lines) correspond to configurations where two or more reflective paths are collinear, thus producing similar ellipses. In this situation, therefore, two measurements yield the same information, thus reducing the robustness of the estimation.

4.3 Disambiguation of TOA Information

The aim of this section is to introduce the methods for disambiguation of TOA information. There inherently exist two major challenges, the permutation and the non-uniqueness problem of TOA matrix. Both challenges are first described and analysed. As a next step, possible solutions are outlined.

4.3.1 Permutation Problem

The permutation problem in the estimated TOA matrix, for the SIMO case, is outlined in the following by considering two illustrative examples. One of the underlying

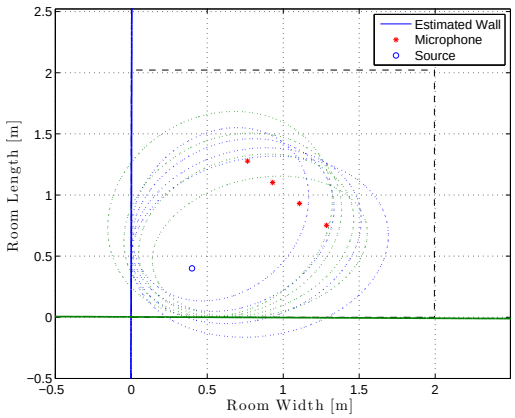


Figure 4.10: TOA matrix permutation problem: Identifying a corner using a microphone array composed of four elements and a single sound source.

$$\left[\begin{array}{c|c|c} \tau_{0,0} & \tau_{0,1} & \tau_{0,2} \\ \tau_{1,0} & \tau_{1,1} & \tau_{1,2} \\ \tau_{2,0} & \tau_{2,1} & \tau_{2,2} \\ \tau_{3,0} & \tau_{3,1} & \tau_{3,2} \end{array}\right]$$

Figure 4.11: Ordering of the TOA matrix for the problem in Figure 4.10.

assumptions used in this work, as developed in Section 3.1.2, was to assume that in $h_i(n)$ the first echo after the direct-path is related to the same reflector. This assumption is satisfied if i) $\mathbf{r}_s$ is sufficiently close to the reflector of interest and ii) microphones are compactly situated in space. We saw that if both conditions apply, then the AIR can

be expressed as

$$h_i(n) = \alpha_{i,0}\delta(n - n_{i,0}) + \alpha_{i,1}\delta(n - n_{i,1}) + \sum_{k=2}^{N} \alpha_{i,k}\delta(n - n_{i,k}).$$

With reference to Figure 4.10, we depict an example, where four microphones and a single source are used to localize the reflectors of a corner in a room. Assuming an otherwise anechoic room, it is straightforward to see, that the two microphones on the upper diagonal will capture an AIR that is composed of the TOAs in the following order: 1) The direct-path propagation from source to microphone; 2) the reflective-path propagation related to the the left wall (blue line); 3) the reflective-path propagation related to the bottom wall (green line). Similarly, the two microphones on the lower diagonal will capture an AIR that is composed of the TOAs in the following order: 1) The direct-path propagation from source to microphone; 2) the reflective-path propagation related to the bottom wall (green line); 3) the reflective-path propagation related to the left wall (blue line). We revisit the notation of the TOA information matrix used in Section 3.1.2, i.e.

$$\boldsymbol{\tau} \triangleq \left[\boldsymbol{\tau}^{\mathrm{DP}} \mid \boldsymbol{\tau}^{\mathrm{RE}}\right].$$

The first column of $\boldsymbol{\tau}$ is always the same, and pertains to the direct-path propagation vector, i.e. $\boldsymbol{\tau}^{\mathrm{DP}}$, from the sound source to each microphone in turn. Since, there are two reflectors present in the acoustic scene, consequently $K = 2$, there are two additional columns in $\boldsymbol{\tau}$, namely the $M \times 2$ matrix

$$\boldsymbol{\tau}^{\mathrm{RE}} = \begin{bmatrix} \tau_{0,1} & \tau_{0,2} \\ \tau_{1,1} & \tau_{1,2} \\ \vdots & \vdots \\ \tau_{M-1,1} & \tau_{M-1,2} \end{bmatrix}.$$

As illustrated in Figure 4.11, there exists an ordering problem along the column dimension of $\boldsymbol{\tau}$. Since $n_{i,0} < n_{i,1} < n_{i,2}$, we note that each row in the TOA information matrix is ordered in time. Because of the relative spatial displacement of each sensor in the array, we can therefore not guarantee that the TOA sequence in each row (i.e. for each microphone) matches the ordering of all other rows. The situation becomes even more complicated if we consider more reflectors, such as all four reflectors in a rectangular room depicted in Figure 4.12, i.e. $K = 4$. In this case, we also violate condition ii), stating that microphones be compactly organized in space. The column-wise

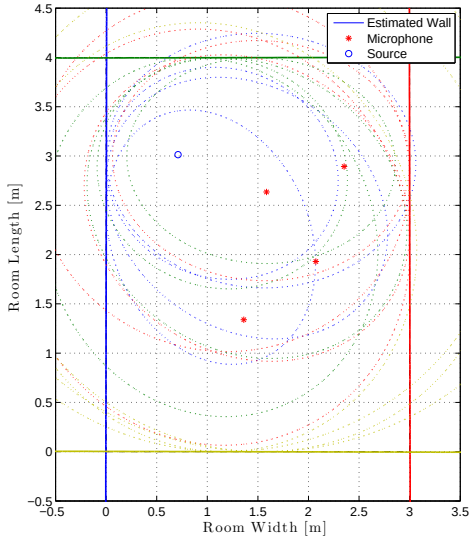


Figure 4.12: TOA matrix permutation problem: Identifying all four reflectors in a rectangular room, measuring 3 × 4 m, using four randomly placed microphones and a single sound source. Reflectors and their corresponding ellipses are drawn in the same color.

$$\left[\begin{array}{c|cccc} \tau_{0,0} & \tau_{0,1} & \tau_{0,2} & \tau_{0,3} & \tau_{0,4} \\ \tau_{1,0} & \tau_{1,1} & \tau_{1,2} & \tau_{1,3} & \tau_{1,4} \\ \tau_{2,0} & \tau_{2,1} & \tau_{2,2} & \tau_{2,3} & \tau_{2,4} \\ \tau_{3,0} & \tau_{3,1} & \tau_{3,2} & \tau_{3,3} & \tau_{3,4} \end{array}\right]$$

Figure 4.13: Ordering of the TOA matrix for the problem in Figure 4.12.

ordering problem for each row in $\boldsymbol{\tau}$ becomes even more complicated, as illustrated in Figure 4.13. Estimating the correct permutation $\boldsymbol{\Gamma}_\pi$ that re-orders the elements in the TOA information matrix, such that every column corresponds to the same reflector, becomes a challenging problem.

4.3.2 Non-uniqueness Problem

There exist some methods in the scientific literature that address the disambiguation of TOA information, such as [110, 154], from which we can take inspiration. These are discussed in greater detail in Chapter 6. Let us note for the moment, that a single room impulse response (RIR) (i.e. a 1-D function of time), uniquely describes the geometry of a planar polygonal room, if we consider both first-order and second-order reflections

[110]. The RIR is therefore a unique signature of a room. This is indeed the case for the assumptions considered in [110], stating that a colocated source and receiver (i.e. the Single-Input Single-Output (SISO) case) are used to identify a unique AIR, i.e. $h_0(n)$, which contains all first and second-order generation delays that provide a mapping between the geometry of a polygonal room and $h_0(n)$.

While the method outlined in [110] approaches the problem of geometric inference in a mathematically elegant way, we note the following important fallacy. Identifying TOAs that are related to all first and second-order reflections is straightforward to obtain in simulation, but often impossible in real reverberant environments. This is mainly due to the fact that a sound source, such as a loudspeaker, does not exhibit an ideal omnidirectional directivity pattern but also because of other effects such as occlusion, non-ideal reflectivity of the building materials and interference. Even more importantly, consider the following scenario: A colocated omnidirectional source and microphone emit and capture a source signal, that is used to identify the RIR, in a square room defined in a 2-D plane. Consider, for illustration purposes, that both source and microphone are located in the centre of this square room. The reflective-path TOAs coincide for all four reflective boundaries of the room. In other words, the RIR only contains a single peak, that is the superposition of all reflective-path peaks. There are many cases, in which peaks in the AIRs seemingly overlap, i.e. one cannot distinguish between them. Note, that this is not only the case for a colocated source and microphone arrangement.

In addition to the permutation problem outlined in the beginning of this section, we therefore also note a non-uniqueness problem in the peaks of the TOA information matrix. In other words, for a convex planar K-polygonal room, we define the set $\mathbf{\Gamma}_0 = \{\tau_{0,j}\}_{1\leq j\leq K}$, that contains first-order TOA information related to the K reflectors for a single receiver ($i = 0$). We denote by $\hat{\mathbf{\Gamma}}_0 = \{\hat{\tau}_{0,j}\}_{1\leq j\leq K}$, the set that contains the estimated TOAs, as identified from the RIR. Evidently it follows that $\hat{\mathbf{\Gamma}}_0 \subseteq \mathbf{\Gamma}_0$. However, based on the illustrative example given earlier, it is possible that $\hat{\mathbf{\Gamma}}_0 \subset \mathbf{\Gamma}_0$, i.e. the inseparability of certain peaks in the RIR lead to $\hat{\mathbf{\Gamma}}_0$ being a proper subset of $\mathbf{\Gamma}_0$.

The disambiguations of TOA information proceeds as follows. We note that there exists both a permutation and non-uniqueness problem in the TOA information matrix. We address these two challenges by considering the more readily available delays related

to first-order reflections and do not, for the moment, include second and higher-order reflections. The motivation for this arises from the fact that we have so far outlined a framework that keeps a priori assumptions on the source signal to a minimum, viz., i) AIRs can be obtained from either supervised or unsupervised methods; ii) the source signal used for excitation can either be a carefully designed pulse or an unknown signal (such as speech, music, the snapping of fingers or the clapping of hands, etc.); iii) source and receivers do not have to be synchronized and no restrictions on the location of the source signal are imposed. In other words, while we limit our model of the AIRs to direct-path and first-order reflection delays, we can consider multiple measurements, for different source positions, in order to disambiguate the TOA information.

We approach the multi-reflector case by performing an exhaustive search using an iterative version of the common tangent algorithm. We adhere to the SIMO case, and only consider a static sound source. While this approach generally produces the desired results it is computationally expensive.

4.3.3 Exhaustive Search

Consider the SIMO case with K reflectors and a single static sound source [114]. By estimating the TOAs of the first order reflections, there exists a set of $M \times K$ ellipses. If a subset of M ellipses are grouped together, extracted from every channel estimate and associated with a particular reflector, then the line parameters of that particular reflector can be estimated using the notation for the cost function developed in Section 3.3, i.e.

$$J_{\mathrm{e}}\left(\mathbf{l}, \left\{\mathbf{C}_{i,k}^{*}\right\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} \left\|\mathbf{l}^{T}\, \mathbf{C}_{i,k}^{*}\, \mathbf{l}\right\|^{2},$$

where $M \geq 3$ and $\mathbf{C}_{i,k}^{*} = \det\left(\mathbf{C}_{i,k}\right)\mathbf{C}_{i,k}^{-1}$. The three unknown line parameters can be estimated by minimizing the cost function

$$\hat{\mathbf{l}}_{k} = \arg\min_{\mathbf{l}} J_{\mathrm{e}}\left(\mathbf{l}, \left\{\mathbf{C}_{i,k}^{*}\right\}_{i=0}^{M-1}\right). \tag{4.25}$$

There is a unique set of M ellipses for every kth reflector. However, prior knowledge is needed to correctly group together related ellipses from the $\frac{(K\cdot M)!}{M!(K\cdot M-M)!}$combinations of possible ellipses. Assuming that the channel estimates only provide TOAs due to first-order reflections, then for a rectangular room in 2-D we can expect $K = 4$ distinct reflective path TOAs in each channel to construct a total of $4 \times M$ ellipses. The

problem then lies in finding the correct M ellipses corresponding to every kth reflector. Exhaustive computation of all combinations of groups of M to find the K optimal line parameters is impractical when M is large. Additionally, if measurement errors are introduced in the system, suboptimal solutions may give erroneous inference results. Under noisy conditions a group of M randomly selected ellipses from the set of $\frac{(K\cdot M)!}{M!(K\cdot M-M)!}$ combinations might produce a better minimum from the cost function than the designated group associated to the kth specific reflector. We therefore propose an iterative approach that groups the set of ellipses on a per-reflector basis [114]. Starting with the geometrically closest reflector to the reference microphone, the total search space is iteratively minimized by discarding ellipses associated with already localized reflectors.

The reflective path TOA associated with the closest reflector to the reference microphone ($\mathbf{r}_0$) is described by $\tau_{0,1}$. The ellipse constructed from this TOA needs to be grouped together with all combinations of ellipses due to other source-microphone pairings and their associated TOAs

$$\tau_{i,k},\ i \in \{0, \cdots, M-1\};\ k \in \{1, \cdots, 4\}.$$

This results in $\frac{(4M)!}{M!(4M-M)!}$ possible combinations. The combination with the smallest value for J_{e} is obtained when all ellipses belong to the same reflector. All ellipses associated with that particular reflector can henceforth be discarded from the search space for subsequent iterations. At the next iteration the search space is reduced to the set of $\frac{(3M)!}{M!(3M-M)!}$ different combinations, then $\frac{(2M)!}{M!(2M-M)!}$ and finally until there are only M ellipses left. Table 4.1 shows how the number of total combinations considered decreases for each subsequent reflector. The Iterative Common Tangent Algorithm (ICOTA) operates well under two conditions. The first condition is that the RIRs

Table 4.1: Algorithm table for estimating reflectors in a rectangular room.

Reflector	Total combinations evaluated	$M = 5$
1^{st}	$\frac{(4M)!}{M!(4M-M)!}$	15504
2^{nd}	$\frac{(3M)!}{M!(3M-M)!}$	3003
3^{rd}	$\frac{(2M)!}{M!(2M-M)!}$	252
4^{th}	M	1

Input: Array of $M \cdot K$ ellipses related to M microphones and K reflectors
Output: Array of $3 \cdot K$ line parameters of K reflectors
foreach *Reflector* k^* **do**
  Set $\mathbf{C}^*_{0,k^*}$;
  foreach *Microphone* i **do**
    foreach *Reflective path TOA* k **do**
      if *Any* $\mathbf{C}^*_{i,k},\ i \in \{0, \cdots, M-1\};\ k \in \{1, \cdots, K\}$ *ellipses not discarded* **then**
        Return minimum $\mathbf{l}$ for $J_{\mathrm{e}}\left(\mathbf{l}, \left\{\mathbf{C}^*_{i,k}\right\}_{i=0}^{M-1}\right)$;
      end
      Discard any $\mathbf{C}^*_{i,k},\ i \in \{0, \cdots, M-1\};\ k \in \{1, \cdots, K\}$ ellipses used;
    end
  end
  Return optimal line parameter $\mathbf{l}$ associated with k^*-th reflector;
end

Algorithm 1: Iterative COTA estimation for multiple reflectors.

contain a complete set of TOA estimates (ideally of only first-order reflections). The second condition is that K is small. Although the algorithm can operate when TOA estimates of higher-order reflections are included in the RIRs, it quickly becomes apparent that because of the way the algorithm scales, it is impractical for real-time applications. For example, when $M = 5$ and $K = 4$, the algorithm would need to run approximately 15500 iterations to localize the first reflector. Including second-order reflections increases the number of iterations to approximately 2119000.

We note, that the ICOTA can be modified to account for the possible non-uniqueness of peaks in the AIRs by not discarding previously used data and consequently performing a full exhaustive search. The motivation in doing this lies in the fact that a global optimal solution exists, which can be estimated optimally. The practicality for such approach, depends on the acquisition scenario considered.

4.4 Improving Robustness using the Hough Transform

In this paragraph the processing of the set of candidate points $\mathbf{P} = [\mathbf{p}_0\, \mathbf{p}_1\, \cdots\, \mathbf{p}_P]$ by means of the Hough transform is considered with the purpose of refining the initial re-

flector estimate $\mathbf{l}$. The Hough transform can be used for estimating the parameters of a shape from its boundary points [142]. It considers the following normal parametrization

$$\rho = x\ \cos\theta + y\ \sin\theta, \tag{4.26}$$

which specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. A point in the cartesian space maps to a sinusoid in the Hough parameter space that corresponds to all the lines passing through it. Conversely, points in the parameter space are transformed into lines in the Cartesian coordinate space. Given two points lying on a line with parameters (ρ, θ), in the Hough parameter space the sinusoids corresponding to these two points intersect at ρ, θ. Therefore, given the points $\mathbf{p}_j$ in the coordinate space, the parameters of a line corresponding to the best-fit of $\mathbf{P}$ can be found. Let $\rho \in \mathbb{R}^+$ and $\theta \in [0, \pi]$. For each point $[x_j\ y_j]^T$ we map

$$\hat{\rho} = x_j\ \cos\hat{\theta} + y_j\ \sin\hat{\theta}. \tag{4.27}$$

The results are stored in an accumulator $\mathcal{A}$, initially set to zero, which is incremented at every step such that:

$$\mathcal{A}\left(\hat{\rho}, \hat{\theta}\right) = \mathcal{A}\left(\hat{\rho}, \hat{\theta}\right) + 1. \tag{4.28}$$

The position of the largest maximum of the accumulator given by

$$\left[\hat{\theta}_{\max}, \hat{\rho}_{\max}\right] = \arg\max\left\{\mathcal{A}\left(\rho, \theta\right)\right\}, \tag{4.29}$$

is then picked, which finally leads to the line parameters of the best-fit:

$$\hat{\mathbf{l}}_{\mathrm{H}} = [\cos(\hat{\theta}_{\max})\ \sin(\hat{\theta}_{\max})\ (-\hat{\rho}_{\max})]^T. \tag{4.30}$$

By taking repeated measurements of TOAs, using a source that is placed at different locations in the acoustic scene, it is possible to append additional data points to $\mathbf{P}$ for a single reflector. True solutions will cluster around the same point in the Hough space, while outliers will receive fewer votes in the accumulator space. There are many robust evaluators available that dynamically remove contributions of backgrounds and analyze voting patterns around peaks in the accumulator space [155]. However, when considering a single reflector in the Hough space, it is often sufficient in practice to estimate the single most voted bin to obtain $\hat{\mathbf{l}}_{\mathrm{H}}$. By computing local centers of energy and discarding outliers in the Hough space, more accurate reflector results can be obtained even when TOA measurements are affected by noise. It will be shown in the following Section how this increases the robustness to noise in the TOA information.

4.5 Experimental Verification

The performance of the proposed inference algorithm for multiple reflectors in 2-D is evaluated through simulations and by using measurements made in real conference rooms. Additionally, a robustness analysis, where a single reflector is considered, highlights the advantages gained by using the Hough transform parametrization. Experiments 1–4 consider simulated impulse responses under a variety of different assumptions and in different scenarios. Specifically, Experiments 1–3 consider the practically noise-less case (i.e. the transfer function of the measurement channel is negligible), while Experiment 4 considers blindly-identified AIRs. In Experiments 5–7, impulse responses are obtained from real-world measurements.

4.5.1 Evaluation Criteria

Given reference source location $\mathbf{r}_\mathrm{s}$ and estimated source location $\hat{\mathbf{r}}_\mathrm{s}$, the source localization error is given by the Euclidian distance $\epsilon_\mathrm{s} = \|\hat{\mathbf{r}}_\mathrm{s} - \mathbf{r}_\mathrm{s}\|$. Let $\mathbf{l}$ and $\hat{\mathbf{l}}$ be the true and estimated reflector lines, respectively. From these we can evaluate the distance d from $\mathbf{r}_0$ to a point on each line and the orientation α. The distance can be evaluated by projecting $\mathbf{r}_0$ onto the line such that

$$d = \frac{|l_1 x_0 + l_2 y_0 + l_3|}{\sqrt{l_1^2 + l_2^2}}, \tag{4.31}$$

and the orientation from

$$\alpha = \arctan \frac{l_2}{l_1}. \tag{4.32}$$

The accuracy of the reflector localization is measured using:

- distance error $\epsilon_\mathrm{d} = \left| d - \hat{d} \right|$;
- angular error $\epsilon_\mathrm{a} = |\alpha - \hat{\alpha}|$;
- alignment error $\epsilon_\mathrm{l} = \frac{\hat{\mathbf{l}}^T \mathbf{l}}{\|\hat{\mathbf{l}}\| \|\mathbf{l}\|}$, where values closer to 1 indicate the angle between the lines is small.

4.5.2 Simulated AIRs

Simulated AIRs were obtained with the source-image method [35, 139], taking into account that propagation delays may not be an integer number of sample periods. The performance was assessed by averaging the results of 100 Monte Carlo runs (unless specified otherwise). The mean and variance of ϵ_{s}, ϵ_{d}, ϵ_{a} and ϵ_{l} were calculated considering all located reflectors and individual reflectors ranked in order of error. In some cases not all reflectors are identified with the same degree of accuracy; ranking the error in this way provides insight into the distribution of errors as a function of the number of identified reflectors.

We consider the SIMO case, such that for each Monte Carlo run the sound source ($\mathbf{r}_{\mathrm{s}}$) was randomly placed at a single static position inside the room. The microphone positions ($\mathbf{r}_i$) were picked from a uniform distribution inside the room, constraining the positions to be at a distance of at least 0.5 m from each wall and with each microphone being kept at a minimum distance of 0.5 m from the source. We exclude those cases in which the inference algorithm fails due either to the inseparability of neighbouring peaks in the AIR, if a source position does not uniquely identify one of the four reflectors or if the matrices involved in the source localization, particularly in (5.8), are rank deficient. In other words, when four reflectors are considered in the acoustic environment, then each AIR should exhibit four first-order peaks; otherwise such arrangement is not taken into account. Furthermore, in the latter case when the microphones are arranged as a linear array it might not be possible to estimate the source location, because of the front-back ambiguity [7]. Additionally the simulation was limited to include only first-order reflections. In all simulations, the sampling frequency is set to 44.1 kHz. We consider unsynchronized AIRs, i.e. source position and range need to be estimated as a first step. Source localization was applied as described in Section 4.1 to estimate TOAs from the TDOAs.

Experiment 1

As an initial simulation, we consider a rectangular room in 2-D of dimensions 4×3 m, using $M = 5$ microphones. The iterative method from Section 4.2.1, based on a constrained least-squares estimator [113, 114] is used to localize the reflectors. Disambiguation of TOA information is addressed using the iterative common tangent

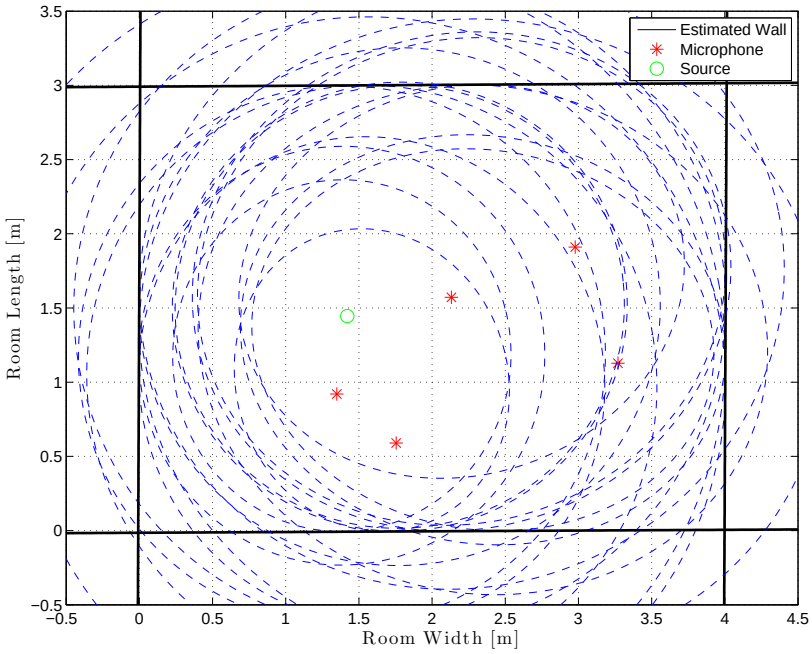


Figure 4.14: Experiment 1: A particular reconstruction result for a room of dimensions 4×3 m.

algorithm. Figure 4.14 shows the estimated reflector lines and the associated ellipses. The results and accuracy of the localization method are shown in Table 4.2. Monte Carlo simulations using the constrained least-squares estimator, along with additional reconstruction examples are given in [114].

Experiment 2

In this experiment we perform Monte Carlo simulations for random source and receiver placement in a rectangular room of random dimensions of width and height ($X \times Y$), with $X \in [3, 5]$ m and $Y \in [4, 6]$ m, using $M = 4$ microphones. The reflector localization is based on the analytic solution of the minimization problem. The Hough transform correction approach is employed for disambiguation of TOA information and quantization. The source localization accuracy is depicted in Table 4.3. We note at this point, that the precision of the source localization plays only a minor role since the error propagation between $\mathbf{r}_s$ and the TOAs is linearly related. In other words, errors in the source localization propagate linearly through the system affecting each channel

Table 4.2: Experiment 1: Comparison of actual and estimated reflector localization results for the room in Figure 4.14

Wall	$(\alpha\,[^\circ], l_3[m])$	$(\hat{\alpha}\,[^\circ], \hat{l_3}[m])$	ϵ_{l}
1	(90,0)	(90.282, 0.015)	0.999
2	(180,4)	(179.751, 3.998)	1.000
3	(-90,3)	(-89.663, 2.991)	1.000
4	(0,0)	(0.249,0.008)	1.000

Table 4.3: Source localization results for simulated AIRs

Exp.	$\mu(\epsilon_{\mathrm{s}})$ [cm]	$\sigma(\epsilon_{\mathrm{s}})$ [cm]
2	0.92	1.62
3	0.92	1.62

Table 4.4: Experiment 2: Distance and angular error results for simulated AIRs

Walls	$\mu(\epsilon_{\mathrm{d}})$ [cm]	$\sigma(\epsilon_{\mathrm{d}})$ [cm]	$\mu(\epsilon_{\mathrm{a}})$ [°]	$\sigma(\epsilon_{\mathrm{a}})$ [°]
All	**0.926**	**1.169**	**0.175**	**0.351**
Best	0.206	0.210	0.036	0.033
2nd best	0.505	0.295	0.079	0.049
2nd worst	0.884	0.421	0.141	0.080
Worst	2.109	1.756	0.442	0.622

equally. The resulting localization error due to this scaling is manifested as a global offset from the origin of the coordinate system. The distance and angular error for the reflector inference are given in Table 4.4. Averaged across all walls our approach achieves a $\mu(\epsilon_{\mathrm{d}})$ and $\mu(\epsilon_{\mathrm{a}})$ of around one cm and less than half degree respectively.

Experiment 3

We conclude our analysis of the reflector localization accuracy by comparing the iterative least-squares estimator with the closed-form estimator. Disambiguation of TOA information is achieved by means of the ICOTA in the former case, and using the Hough

Table 4.5: Experiment 3: Iterative LS estimator alignment error

Walls	$\mu(\epsilon_\mathrm{l})$	$\sigma(\epsilon_\mathrm{l})$
All	**0.996**	**0.055**
Best	1.000	0.000
2nd best	1.000	0.005
2nd worst	0.996	0.033
Worst	0.987	0.105

Table 4.6: Experiment 3: Closed-form estimator alignment error

Walls	$\mu(\epsilon_\mathrm{l})$	$\sigma(\epsilon_\mathrm{l})$
All	**1.000**	**0.000**
Best	1.000	0.000
2nd best	1.000	0.000
2nd worst	1.000	0.000
Worst	1.000	0.000

Table 4.7: Experiment 3: Iterative LS estimator distance and angular error

Walls	$\mu(\epsilon_\mathrm{d})$ [cm]	$\sigma(\epsilon_\mathrm{d})$ [cm]	$\mu(\epsilon_\mathrm{a})$ [°]	$\sigma(\epsilon_\mathrm{a})$ [°]
All	**3.720**	**16.580**	**0.799**	**3.258**
Best	0.290	0.230	0.046	0.074
2nd best	1.390	6.130	0.264	1.606
2nd worst	3.820	12.300	0.956	3.544
Worst	9.400	29.490	1.931	5.048

Table 4.8: Experiment 3: Closed-form estimator distance and angular error

Walls	$\mu(\epsilon_\mathrm{d})$ [cm]	$\sigma(\epsilon_\mathrm{d})$ [cm]	$\mu(\epsilon_\mathrm{a})$ [°]	$\sigma(\epsilon_\mathrm{a})$ [°]
All	**0.926**	**1.169**	**0.215**	**0.426**
Best	0.206	0.210	0.034	0.030
2nd best	0.505	0.295	0.091	0.057
2nd worst	0.884	0.421	0.179	0.138
Worst	2.109	1.756	0.555	0.737

space approach in the latter. The source localization accuracy is shown in Table 4.3. The results of the alignment, and distance and angular error are given in Tables 4.5–4.6 and Tables 4.7–4.8 respectively. In terms of alignment, the four walls are accurately localized using the ICOTA [114]. The closed-form estimator on the other hand yields a perfect alignment in all cases, $\mu(\epsilon_\mathrm{l}) = 1$. In terms of distance and angular error the closed-form estimator outperforms the original line estimator with lower mean and

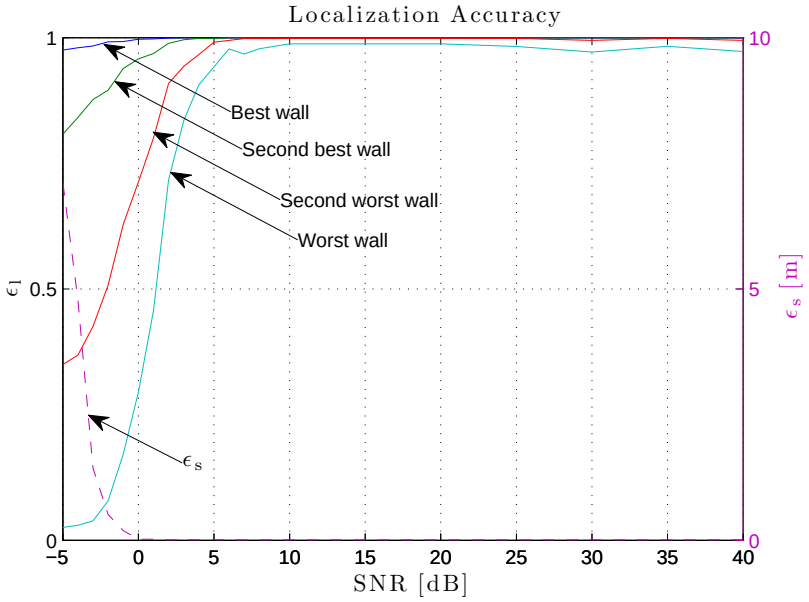


Figure 4.15: Blind system identification experiment: Alignment error (ranked by accuracy), shown as the left vertical axis, and source localization accuracy, shown as the right vertical axis, as a function of SNR.

considerably smaller variance. Averaged across all walls the analytic line estimator presented in this work achieves a $\mu(\epsilon_{\mathrm{d}})$ and $\mu(\epsilon_{\mathrm{a}})$ of less than one cm and one degree respectively.

Experiment 4

We consider in this experiment, blindly-identified AIRs in order to analyse the behaviour of the reflector inference algorithm under varying SNR conditions. The simulated AIRs are convolved with a WGN signal of duration 5 s and the channels estimated with the RNMCFLMS algorithm [128] with parameters $\rho = 0.2,\ \lambda = 0.98$. In order to prevent overmodeling of the AIRs, the effective length of the channels is estimated by $\max(\tau_{i,k}) - \min(\tau_{i,k})$, where ground-truth $\tau_{i,k}$ are used; the observed signals $x_i(n)$ are otherwise the only signals available to the BSI algorithm. Uncorrelated sensor noise is added to give SNR $\{-5, -4, \ldots, 40\}$ dB, providing insight into the behaviour of the inference algorithm with different levels of noise. The BSI numerical simulations were conducted in 200 Monte Carlo runs.

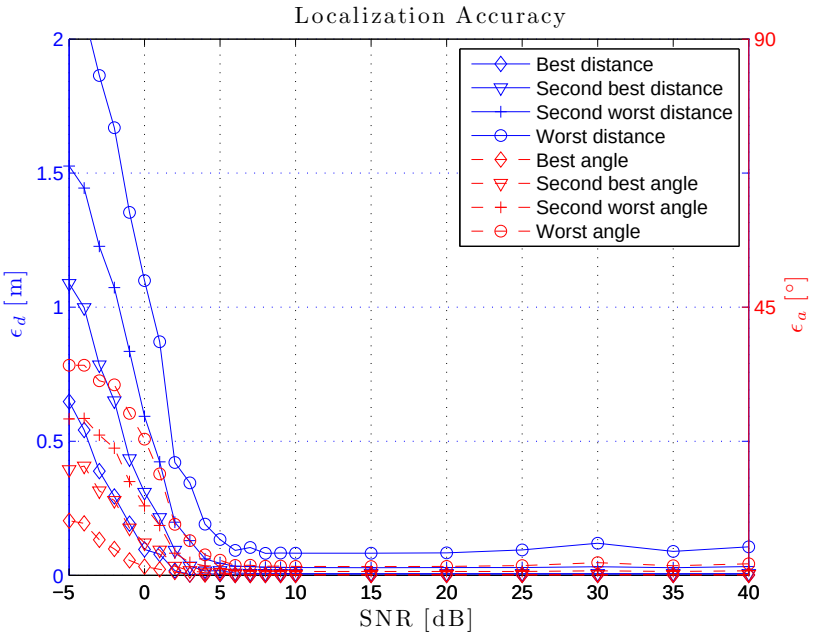


Figure 4.16: Blind system identification experiment: Distance, shown as the left vertical axis, and angular reflector localization error (ranked by accuracy), shown as the right vertical axis, as a function of SNR.

The results of Experiment 4 are shown in Figs. 4.15–4.17. Reliable localization of all four walls can be achieved at input SNR values of 10 dB or greater. Errors begin to occur at SNR < 10 dB although on average at least two walls can be identified at SNR$= -5$ dB ($\epsilon_1 \geq 0.8$). Existing applications of BSI, such as dereverberation by channel equalization [2], usually require significantly high SNRs in order to be effective as it is required that all taps be modeled well. The BSI problem is relaxed in the case of inference as only the time of the early reflections, which tend to be of high amplitude compared with neighbouring taps, need be known. Further work into BSI in the context of inference is expected to exploit this relaxed requirement for the identified channels.

4.5.3 Real-World Results

The simulated experiments represent idealized environments in which the transfer function of the measurement apparatus is negligible and the floor and ceiling are perfectly absorbing. In the case of real-world measurements, geometric inference is a much more challenging problem. Two experiments were devised to demonstrate the applicability of

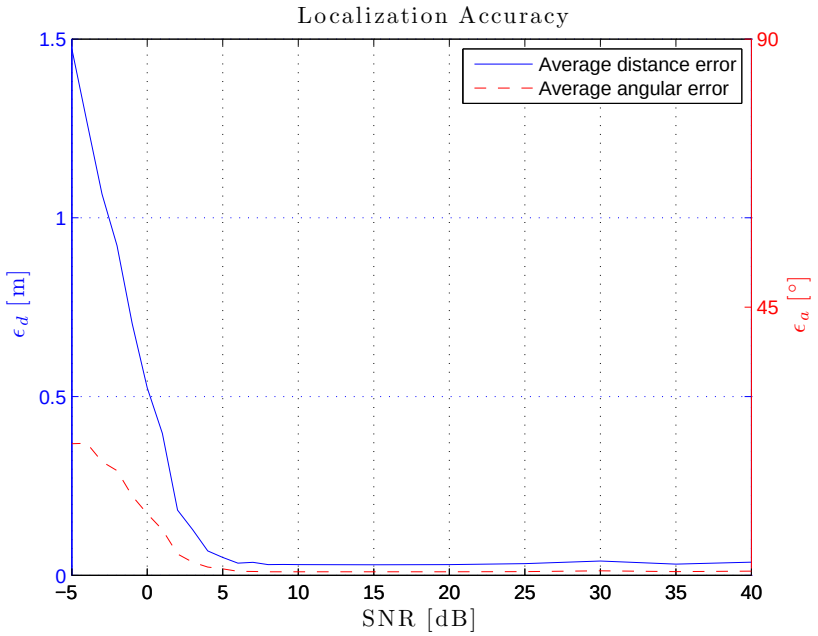


Figure 4.17: Blind system identification experiment: Average distance, shown as the left vertical axis, and angular reflector localization error (for all walls combined), shown as the right vertical axis, as a function of SNR.

the reflector localization algorithm in a real-world environment. As a first example, the corner of a rectangular conference room is estimated at four different source locations using a linear microphone array. As a next step, all four walls of different conference room are estimated by combining up to 16 independent source position measurements in the Hough parameter space.

Experiment 5

The red $*$ in Fig. 4.18 show a microphone array consisting of four DPA 4063 microphones spaced by 16 cm in a linear configuration. A fifth microphone was placed in the centre of the array, displaced 5 cm perpendicular to the array line on the horizontal plane. This aids in resolving the front-back ambiguity. The array was placed in the corner of a 4.77 $\times$ 5.92 m room at a height of 1.2 m parallel to one wall with the reference microphone ($i = 0$) displaced (1.0, 0.5) m relative to the corner. In the coordinate system of Figure 4.18, the corner of the room lies at (0.0, 0.0) m and the walls extend in the positive x and y directions. A Genelec 8030A loudspeaker was placed sequentially

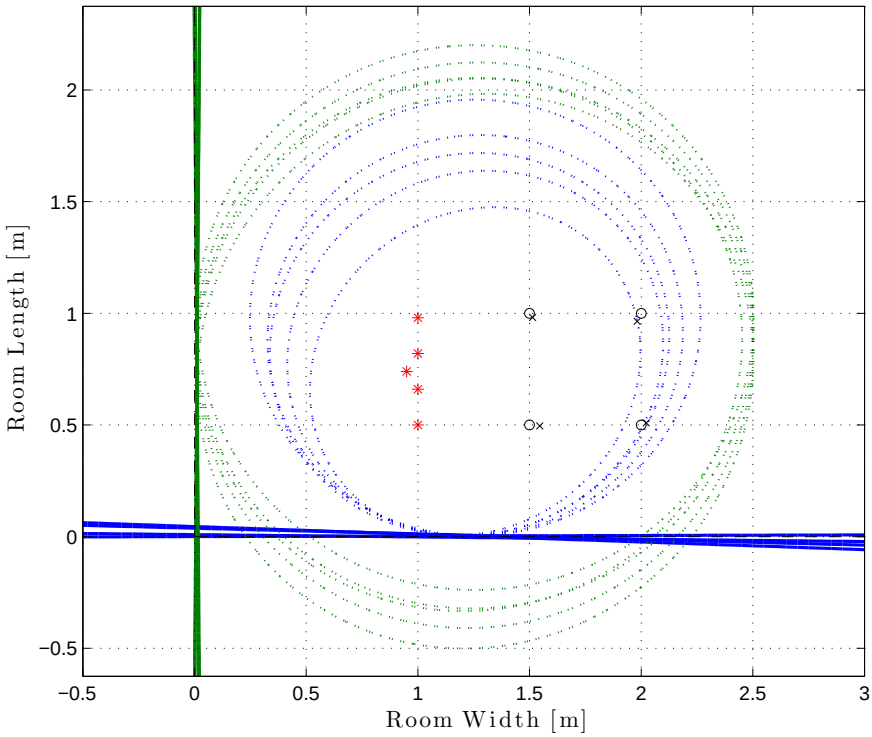


Figure 4.18: Inference in a real-world environment. Five microphones (red '$*$') and a single loudspeaker placed at 4 sequential points (black '$\circ$') were used to estimate two perpendicular reflectors passing through (0, 0) m. The elliptical constraints are shown pertaining to source location (1.5, 1) m. The estimated source locations are depicted '$\times$' and the estimated reflectors for all 4 source positions are overlaid around $x = 0$ and $y = 0$.

at four positions: (2.0, 0.5), (1.5, 0.5), (2.0, 1.0) and (1.5, 1.0) m as marked by black $\circ$. The array and source locations were placed square to the reflectors simply to aid their placement within the room as this is not expected to have an impact upon the inference algorithm. At each source location, the AIRs were measured with the MLS method [140]. No effort was made to synchronize the recorded signals with the input stimulus.

The AIRs in Experiment 5 were analysed assuming unsynchronized AIRs and an unknown source location. In each case, the three largest peaks in the AIRs were assumed to correspond to the direct-path signal and 1st-order reflections due to the nearby walls. The black $\times$ in Figure 4.18 show the estimated locations of the sources,

exhibiting a mean error of 3.4 cm. The thick horizontal and vertical lines in Figure 4.18 depict the estimated reflectors overlaid for all four source locations. For clarity, the ellipses shown pertain to an arbitrarily selected example of source location (1.5, 1.0) m only. The inference results are summarized in Table 4.9 and show that the reflectors were localized to within 2.8 cm and 0.7 degrees. The low inference error suggests that the source location error was due in part to the difficulty in estimating a point source from a loudspeaker by physical measurement. These preliminary results suggests that the proposed algorithm is suitable for real-world measurements.

Table 4.9: Experiment 5: Reflector localization results with real-world data

Walls	$\mu(\epsilon_{\mathrm{l}})$	$\mu(\epsilon_{\mathrm{d}})$ [cm]	$\mu(\epsilon_{\mathrm{a}})$ [°]
All	**0.999**	**2.753**	**0.714**
Best	0.999	1.236	0.326
2nd Best	0.999	4.270	1.103

Experiment 6

An experiment was devised in a small conference a room measuring $3.31 \times 3.58 \times 3.00$ m, with concrete walls and two flush-mounted wooden doors in the south and east walls. A microphone array consisting of four microphones spaced by 0.5 m in a '+' configuration and a fifth placed in the centre was positioned at (1.75,1.50) m from the south-west corner. A Genelec 8030A loudspeaker was positioned around the array in 16 equiangle positions at a range of 1 m from the array centre, ensuring that the loudspeaker was always faced towards the array. The loudspeaker positions used in this experiment are similar to those used in a 2-D wave field synthesis array. The microphone signals were sampled at 44.1 kHz. At each position, the acoustic impulse response between the source and microphone array was estimated using the MLS method [140]. The recorded signals were not synchronized with the input stimulus. The line estimates were combined using the Hough transform and the parameters corresponding to the top four bins were used to estimate the bounding line reflectors.

We proceeded to evaluate the improvement in localization accuracy when an increasing number of source positions is employed. First, the sound source was arranged

Table 4.10: Experiment 6: Reflector localization results with real-world data

	$N = 4$		$N = 8$		$N = 12$		$N = 16$	
Wall	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]
All	**2.358**	**0.812**	**1.580**	**0.603**	**1.025**	**0.309**	**0.853**	**0.205**
North	2.240	0.751	2.240	0.751	0.700	0.253	0.700	0.253
East	1.780	0.720	0.850	0.751	0.610	0.504	0.500	0.252
South	2.960	0.877	1.780	0.419	1.680	0.226	1.560	0.126
West	2.450	0.899	1.450	0.490	1.111	0.252	0.650	0.188

in a '+' configuration, i.e. displaced on the north, west, south and east directions with respect to the array. As a next step, four further source positions are considered at a rotation of 45°, i.e. including measurements coming from the north-west, south-west, south-east, north-east. Finally, two further rotations of $+22.5°$ and $-22.5°$ yield results for 12 and 16 source positions. Table 4.10 shows the localization accuracy for each of the walls along with the average accuracy for all the four configurations described above. Notice that the accuracy improves as the number of sources increases. Even for the case of 4 source positions an error of only a few centimeters is observed, which is suitable for many application scenarios. Using 16 source positions, effectively mimicking a wave field synthesis array, the localization accuracy approaches the limits of the hand-measured ground truths. Localization results for the 16 sources case are shown in Figure 6.1. The error between the intended and estimated positions is due, in part, to the manual measurement of the position of the loudspeaker. This is not problematic as the system makes no prior assumptions about the source location. The Hough data points, marked as '+', lie very close to the room boundaries and are well-fitted by the estimated line reflectors. Some erroneous data points are due to the source positions near multiples of 45° in which no single reflector is dominant; they are however treated as outliers by the algorithm and do not affect the estimated reflectors. Reflections from the walls were always dominant over those arising from the floor and ceiling as they are less reflective than the walls, i.e. any reflections from the floor and the ceiling were negligible because of their construction material.

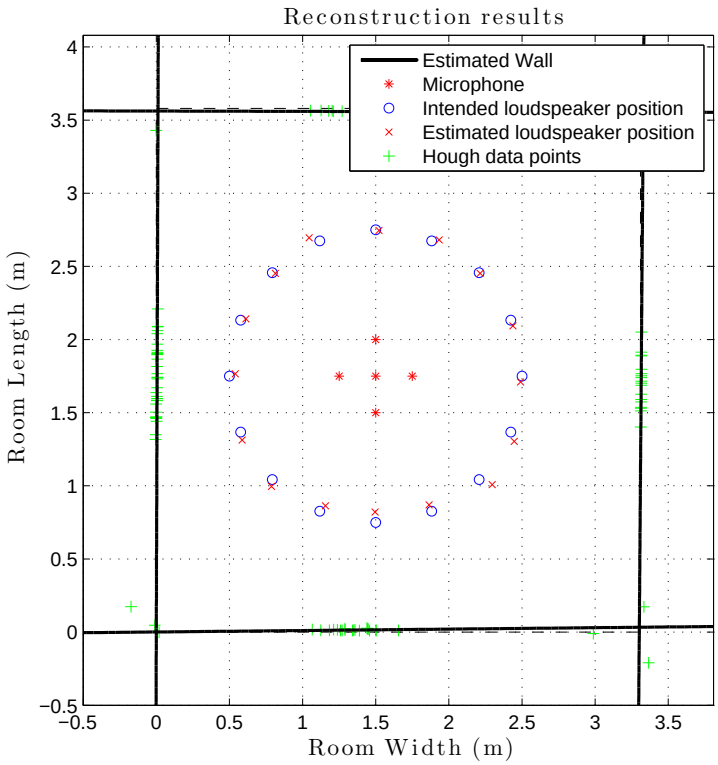


Figure 4.19: Room inference results using a microphone array, placed centrally in a small conference room, capturing a MLS sequence from 16 source positions in turn.

Experiment 7

In this experiment the Hough transform based inference approach is validated in a medium sized conference room. The crux in this experiment lies in the fact that the room impulse response is identified using an unsupervised approach with an unknown signal: a test person is snapping their fingers while moving around the room.

The finger-snaps were captured with a microphone array that consisted of four DPA 4063 microphones spaced by 16 cm in a linear configuration and a fifth microphone that was placed in the centre of the array, displaced 5 cm perpendicular to the array line on the horizontal plane. The array was placed in the corner of the room at a height of 1.2 m facing the south wall with the leftmost microphone displaced (0.61, 0.95) m relative to the corner. The arrangement is shown in Figure 4.20.

A test person performs a "random walk" in the vicinity of the microphone ar-

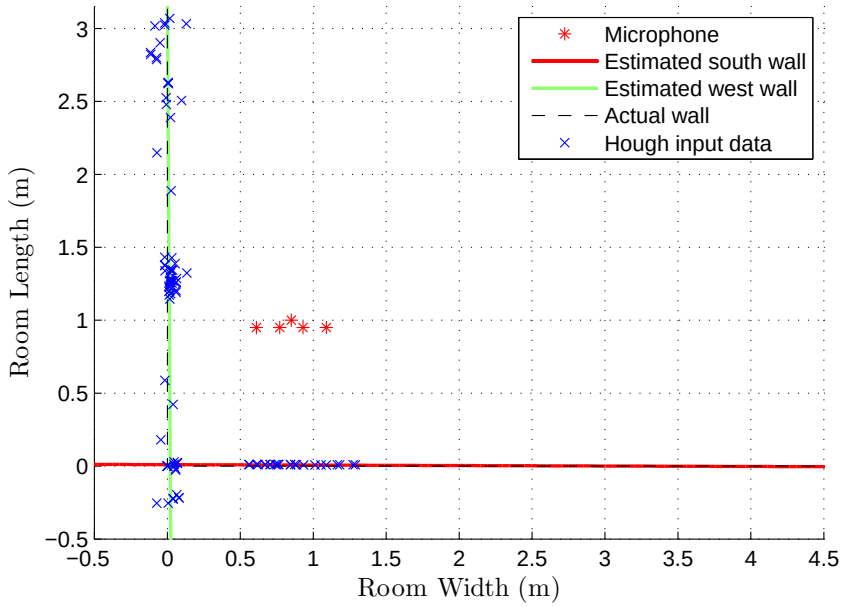


Figure 4.20: Localization of the corner walls in a medium sized conference room. The arrows depicted the approximate trajectory of the "random walk" of a test person that is snapping their fingers at different spatial locations.

ray snapping their fingers at different spatial positions. The direct and the dominant reflective-path TOAs were extracted from each channel at every position where a finger-snap was localized and captured. At each step, the line parameters of the single dominant reflector were found using the closed-form estimator. From this localized reflector, Hough input data points were constructed. These input data points were accumulated for every finger-snap captured. After post-processing the data in the Hough parameter space the largest two local maxima were picked. Note that in this example, the number of maxima are set in a priori way - this is because we have purposely placed the microphone array close to the corner of the room. From these two maxima the line parameters of the actual reflectors are estimated robustly.

Given more measurements, or *evidence*, it can be seen that the accuracy of the reflector localization improves. The Hough input points cluster around the true reflector location, as seen in Figure 4.20. Data points that deviate from the true reflector

Table 4.11: Experiment 7: Reflector localization results with real-world data

	5 Measurements		15 Measurements	
Wall	ϵ_{d} [cm]	ϵ_{a} [°]	ϵ_{d} [cm]	ϵ_{a} [°]
All	**2.765**	**1.714**	**1.530**	**0.857**
South	1.020	0.900	0.640	0.453
West	4.510	2.528	2.420	1.261

location can be suppressed using thresholding methods in the Hough space and by applying *thinning* to the isolated clusters of bright spots in the accumulator array image. Table 4.11 shows the reflector localization accuracy when 5 and 15 finger-snaps are captured.

Interestingly, not only has this real-world experiment shown that a room corner can be localized accurately using an unknown source signal (finger-snap), furthermore measurements at random source positions in the proximity of the microphone array lead to the creation of "hot-spots" in the Hough parameter space. Theoretically, this could lead to the mapping of more reflectors and more complex geometries (i.e. not restricted to rectangular rooms). However, restricting the algorithm to probe for only dominant "hot-spots" (in this case two) yields important information about the local environment of the measurement apparatus. For example, the microphone array could be a teleconferencing system. Knowing the location of the two dominant reflectors close to the teleconferencing system, and improving their location estimate over time, can be of significant advantage for acoustic echo cancellation algorithms. The speaker would merely walk around the room and talk and the system could automatically improve the localization accuracy of the offending reflectors (and possibly reduce their impact).

4.5.4 Robustness Analysis

In order to study the robustness of our method with respect to noise, additional white noise was added to the TOA estimates of a single reflector:

$$n_{i,1}^{\star} = n_{i,1} + \xi, \tag{4.33}$$

where ξ is zero-mean Gaussian noise with standard deviation variable between 0 and 5 samples.

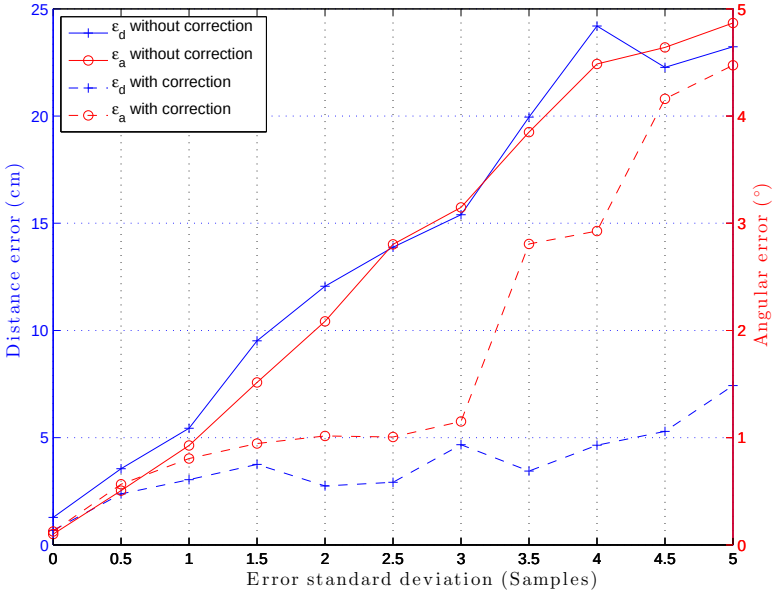


Figure 4.21: Average distance, shown as the left vertical axis, and angular reflector localization error, shown as the right vertical axis, for a single reflector using five linearly arranged source positions, as a function of additive noise to the TOA estimates for the estimate without the Hough transform (without correction) and with the Hough transform (with correction).

Two arrangements of source positions were considered: a linear and circular arrangement. In the first case the source was placed 0.5 m behind the centre of the microphone array (with respect to the wall) and moved at five equidistant intervals between $[-10, 10]$ cm along the length of the room. In the second case the source was moved on a half circle of diameter 1 m from the centre of the array at five equiangle positions between $[-270°, 90°]$.

The performance was assessed by averaging the results of 50 Monte Carlo runs. In each run five source positions were used. For each source position the line parameters of the reflector were calculated using the initial estimate (4.13) and the Hough data points. At the end of each run the average error of the COmmon TAngent (COTA) method was computed along with the best fit, obtained from the analysis of the Hough parameter space, of the five repetitions combined.

The results for both arrangements are shown in Figures 4.21–4.22 respectively. In

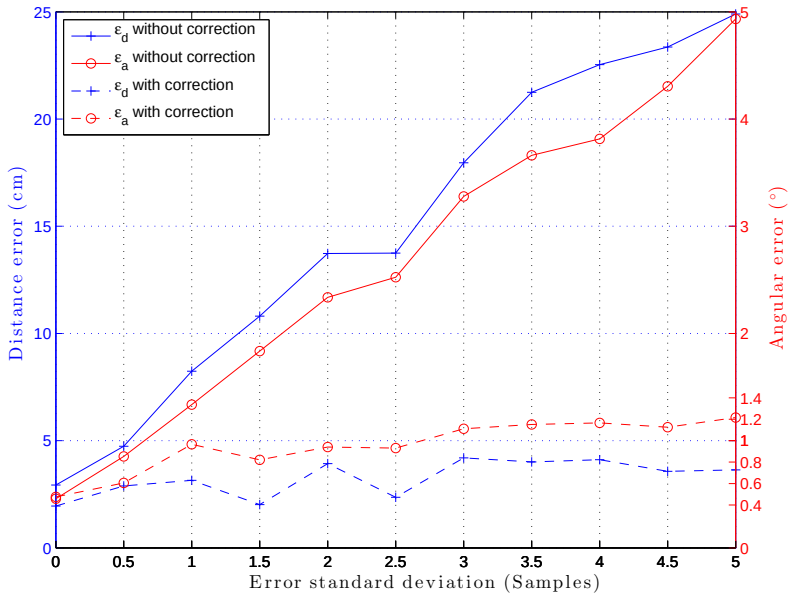


Figure 4.22: Average distance, shown as the left vertical axis, and angular reflector localization error, shown as the right vertical axis, for a single reflector using five circularly arranged source positions, as a function of additive noise to the TOA estimates for the estimate without the Hough transform (without correction) and with the Hough transform (with correction).

both cases it is observed that the robustness to errors in the TOAs is improved by clustering multiple measurements. When the source positions are close to each other, i.e. yielding a low spatial variation, such as in Fig. 4.21, we see that the angular error (and also to some extent the distance error) increases rapidly with added errors to the TOA information. Consequently, enough spatial variation should be introduced in the positioning of the source if high accuracy is desired, especially in adverse conditions, i.e. when TOA estimates are noisy.

Conclusion

The concepts developed in this chapter outline the problem of reflector line parameter estimation from AIRs. Specifically, inference of the geometry of an acoustic environment in 2- D from AIR estimates has been considered in this chapter. Peaks in the AIRs correspond to the TDOAs related to the dominant reflections in a room, from

which the location of the source relative to the receivers can be estimated. The TOA corresponding to each peak can then be used in conjunction with the relative source and receiver locations to parameterize an ellipse that describes the locus of possible reflector locations. The common tangent between multiple ellipses corresponds to the location of a particular reflector. An algorithm has been proposed that automatically locates multiple reflectors in a 2-D plane from estimates of the AIR. Monte Carlo simulations reveal that the proposed method works reliably even when the source location is unknown and the AIRs are unsynchronized. Further simulations show that, by using the Hough transform and taking repeated measurements at different source positions, the robustness to noise in the TOA information can be improved. Real-world measurements show that the proposed technique provides reliable results in a practical setting.

5

Room Geometry Estimation in Three-Dimensions

Introduction

WE extend the reflector localization framework outlined in Chapter 4 to the three-dimensional (3-D) case. Instead of dealing with line reflectors the aim is now to estimate the parameters of a reflector plane. As a first step, the notation of the system model is extended to the 3-D case where multiple acoustic impulse responses (AIRs) are captured at different spatial positions. Since we do not impose any a priori assumptions on the system identification process, whether it is supervised or unsupervised, the source localization algorithm, enabling the identification of time-of-arrivals (TOAs) from unsynchronized measurements, is outlined in 3-D. We then introduce the planar reflector estimation procedure.

We consider two methods. First, we introduce a method that estimates planar reflector parameters by combining multiple orthogonal line parameters. More specifically, we adopt a 3-D array accommodating seven microphones. Microphones are organized in three sub-arrays, each composed of five microphones. All the microphones in a specific sub-array are characterized by the fact that they are co-planar. Each sub-array is devoted to the localization of the portion of reflectors lying on its plane. By intersecting line-reflectors estimated from multiple sub-arrays, the proposed methodology estimates the actual lying plane of each reflector.

Second, the approach in Section 4.2.3, that was proposed in [116], is extended to 3-D

geometries. Additionally, the analytical prediction of the impact of errors on measurements on the reflector localization error is extended to the 3-D case. This research theme is becoming increasingly important, as demonstrated by recent publications (e.g. [156] and [157]) where the authors study the problem of the propagation of error measurements into the estimation process, taking inspiration from the Information Geometry [158].

The remainder of this chapter is organized as follows: Section 5.1 introduces the notation of the system model. Section 5.2 outlines the source localization algorithm in 3-D. The first reflector localization method, that combines multiple line reflectors to estimate the planar reflectors, is introduced in Section 5.3. Section 5.4 presents the constrained least-squares and exact estimator. Finally, Section 5.5 gives experimental results.

Relevant Publications

1. J. Filos, A. Canclini, F. Antonacci, A. Sarti and P. A. Naylor, "Localization of Planar Acoustic Reflectors from the Combination of Linear Estimates," *Proc. European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, August 27 - 31, 2012.

2. A. Canclini, F. Antonacci, J. Filos, A. Sarti and P. A. Naylor, "Exact Localization of Planar Acoustic Reflectors in Three-Dimensional Geometries," *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, September 4 - 6, 2012.

5.1 System Model

We will consider from now on M sensors that are distributed in a 3-D volume at positions $\mathbf{r}_i \triangleq [x_i \ y_i \ z_i]^T, \ i = 0, \dots, M-1$. We assume N source positions and extend our notation by adding the source location index m such that it is given by $\mathbf{r}_{\mathrm{s},m} \triangleq [x_m \ y_m \ z_m]^T, \ m = 1, \dots, N$. The received signal, as formulated in Section 2.1, at the ith sensor and related to the mth source, is then given by the extension of (2.1), i.e.

$$x_{i,m}(n) = h_{i,m} * s_m(n) + b_{i,m}(n), \tag{5.1}$$

which can be written in vector form as

$$\mathbf{x}_{i,m}(n) = \mathbf{H}_{i,m}(n)\mathbf{s}_m(n) + \mathbf{b}_{i,m}(n), \tag{5.2}$$

where

$$\begin{aligned}
\mathbf{x}_{i,m}(n) &= [x_{i,m}(n)\; x_{i,m}(n-1)\; \cdots\; x_{i,m}(n-L+1)]^T, \\
\mathbf{H}_{i,m}(n) &= \begin{bmatrix} h_{i,m,0}(n) & \cdots & h_{i,m,L-1}(n) & \cdots & \cdots & 0 \\ 0 & h_{i,m,0}(n) & \cdots & h_{i,m,L-1}(n) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & h_{i,m,0}(n) & \cdots & h_{i,m,L-1}(n) \end{bmatrix}, \\
\mathbf{s}_m(n) &= [s_m(n)\; s_m(n-1)\; \cdots\; s_m(n-2L+2)]^T, \\
\mathbf{b}_{i,m}(n) &= [b_{i,m}(n)\; b_{i,m}(n-1)\; \cdots\; b_{i,m}(n-L+1)]^T.
\end{aligned}$$

The impulse response of each channel $h_{i,m}(n)$ is now related to a particular source position such that

$$h_{i,m}(n) = \alpha_{i,m,0}\delta(n - n_{i,m,0}) + \alpha_{i,m,1}\delta(n - n_{i,m,1}) + \sum_{k=2}^{N} \alpha_{i,m,k}\delta(n - n_{i,m,k}), \tag{5.3}$$

where $\alpha_{i,m,k}$ is the attenuation along the direct ($k = 0$) or reflective ($k \geq 1$) path for every mth source position. Consequently, we extend the notation for the TOA information matrix using the additional index m, i.e. $\boldsymbol{\tau}_m, m = 1, \cdots, N$, for M microphones, N source positions and K reflectors such that

$$\boldsymbol{\tau}_m = \begin{bmatrix} \tau_{0,m,0} & \tau_{0,m,1} & \tau_{0,m,2} & \cdots & \tau_{0,m,K} \\ \tau_{1,m,0} & \tau_{1,m,1} & \tau_{1,m,2} & \cdots & \tau_{1,m,K} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \tau_{M-1,m,0} & \tau_{M-1,m,1} & \tau_{M-1,m,2} & \cdots & \tau_{M-1,m,K} \end{bmatrix}. \tag{5.4}$$

Note, that we never acquire $\boldsymbol{\tau}_m$ for all source positions at one point in time, i.e. we do not consider the MIMO case, but rather estimate the $M \times K$ TOA matrix multiple times for N source positions. This is why the notation $\boldsymbol{\tau}_m$ is convenient, as we measure the AIRs in the same way as before, but collect $m = 1, \cdots, N$ instances. Consequently, we reflect the above changes to $\tau_{i,j}^{\mathrm{TDOA}}$ which is now defined as the time-difference-of-arrival (TDOA) of the direct-path between the ith and jth microphone for every mth source position such that

$$\tau_{i,m,j}^{\mathrm{TDOA}} = |\tau_{i,m,0} - \tau_{j,m,0}|. \tag{5.5}$$

5.2 Source Localization in 3-D

In this section, we briefly outline the original source localization formulation used (that is defined in 3-D in [80]), that extends the notation developed for the 2-D case in Section 2.3.2. For notational convenience, but without loss of generality, we consider only a single static source, i.e. $\mathbf{r}_{\mathrm{s},m}$ becomes $\mathbf{r}_{\mathrm{s}} \triangleq [x_{\mathrm{s}}\ y_{\mathrm{s}}\ z_{\mathrm{s}}]^T$.

The reference microphone ($i = 0$) is placed at the origin of the new coordinate system, i.e. $[0\ 0\ 0]^T$ and the distances from the origin to the ith microphone and the source are given by

$$R_i \triangleq \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2 + z_i^2}, \quad i = 0, \ldots, M-1,$$
$$R_{\mathrm{s}} \triangleq \|\mathbf{r}_{\mathrm{s}}\| = \sqrt{x_{\mathrm{s}}^2 + y_{\mathrm{s}}^2 + z_{\mathrm{s}}^2}.$$

We observe that the correct source location should be at the intersection of a group of spheres, analogously to the circular criterion in Section 2.3.2. The centre of each sphere is equal to the location of the microphone and the radius of each sphere is related to the source-microphone distance. Therefore, the best estimate of the source location will be the point that yields the shortest distance to those spheres defined by the range differences and the hypothesized source range. The error function is then defined as the difference between the measured and true values, which when writing them in a vector form gives

$$\mathbf{e}(\mathbf{r}_{\mathrm{s}}) = \mathbf{A}\theta - \mathbf{b}, \tag{5.6}$$

where

$$\mathbf{A} \triangleq \left[\mathbf{S} \mid \hat{\mathbf{d}}\right], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_{M-1} & y_{M-1} & z_{M-1} \end{bmatrix},$$
$$\theta \triangleq \begin{bmatrix} x_{\mathrm{s}} \\ y_{\mathrm{s}} \\ z_{\mathrm{s}} \\ R_{\mathrm{s}} \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{1,0}^2 \\ R_2^2 - \hat{d}_{2,0}^2 \\ \vdots \\ R_{M-1}^2 - \hat{d}_{M-1,0}^2 \end{bmatrix}.$$

The corresponding spherical LS criterion is then given by

$$J = \mathbf{e}^T\mathbf{e} = [\mathbf{A}\theta - \mathbf{b}]^T [\mathbf{A}\theta - \mathbf{b}]. \tag{5.7}$$

The solution for θ is given by [80]

$$\hat{\theta}_1 = \mathbf{A}^{\dagger}\mathbf{b}, \tag{5.8}$$

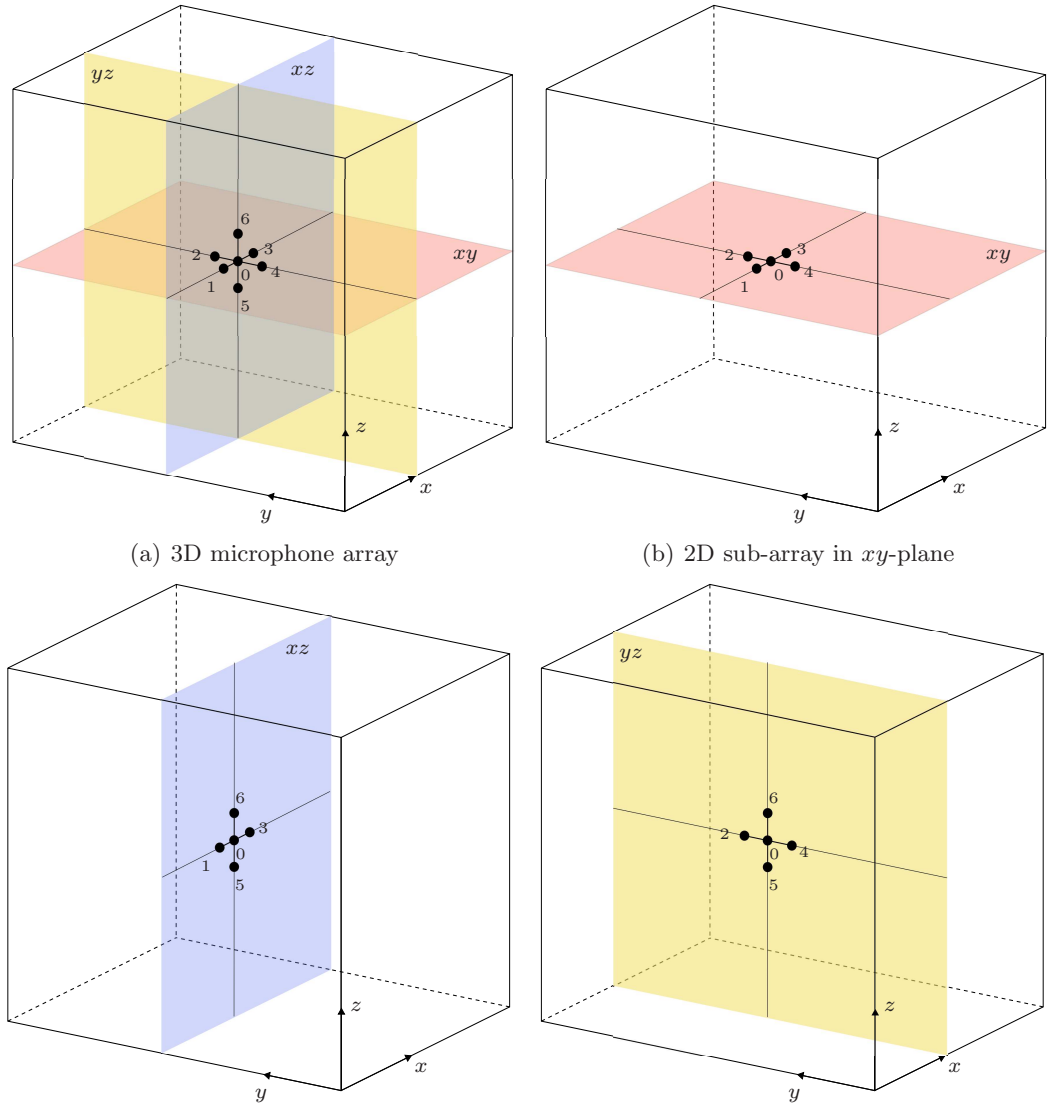
where $(\,\cdot\,)^{\dagger}$ denotes the pseudo-inverse.

5.3 Reflector Plane Localization from Combination of Linear Estimates

Many techniques have appeared in the last few years, which aim at localizing principal reflectors in a room. Relevant examples are [104, 108, 110, 113, 114]. All these techniques, however, specifically address the estimation of 2-D geometries. There are many scenarios where reflections from floor and ceiling are relevant and can affect the accuracy and outcome of the space-time processing. In [117] the authors generalize the approach in [113, 114] to 3-D geometries. In this section we start once again from the approach introduced in [113, 114] but propose a rather different approach to the estimation of simple 3-D geometries, which transforms the localization of planar reflectors into the estimation of multiple linear reflectors. More specifically, we adopt a 3-D array accommodating seven microphones. Microphones are organized in three sub-arrays, each composed of five microphones. All the microphones in a specific sub-array are characterized by the fact that they are co-planar. Each sub-array is devoted to the localization of the portion of reflectors lying on its plane. By intersecting line-reflectors estimated from multiple sub-arrays, the proposed methodology estimates the actual plane of each reflector.

An AIR acquired in an ordinary room can be richly populated with peaks related to reflective paths, only some of which are related to first-order reflections. We consider these first-order echoes as the only acoustic events useful for the localization of reflectors in this scenario, as a simplifying assumption. A preliminary step that selects only the useful acoustic events, i.e. the TOAs related to the direct-path propagation and the first-order reflective paths, is therefore necessary. For this purpose we propose a technique based on the Hough transform. The Hough transform for the detection of reflectors was first introduced in [115]. Based on the assumption that all the cartesian sections of the room are rectangular, we select the reflective paths in the

impulse response which are organized on a rectangular pattern in the Hough parameter space. This rectangle detection technique is inspired by the solution to a similar problem adopted in computer vision [141]. It is worth noticing that this approach can be generalized to convex polygonal rooms [159].



(a) 3D microphone array

(b) 2D sub-array in xy-plane

(c) 2D sub-array in xz-plane

(d) 2D sub-array in yz-plane

Figure 5.1: 7-element microphone array inside a room: Full 3D array (a) and decomposition into three 2D sub-arrays (b)–(d).

5.3.1 Proposed Method

In this section, we aim to obtain the full TOA set $\boldsymbol{\tau}_m, m = 1, \cdots, N$, i.e. the set that allows the identification of all reflectors (i.e. walls) that define the boundaries of the acoustic enclosure, by considering multiple source positions. This is achieved by capturing the AIRs at different source locations. For each source location the multi-channel impulse response, $h_{i,m}$, is identified. The TOA information matrix $\boldsymbol{\tau}_m$ is created by picking the peaks from $h_{i,m}$, that represents the impulse responses from each source ($\mathbf{r}_{\mathrm{s},m}$) to each microphone ($\mathbf{r}_i$) in the array.

Two limitations are considered. First, the number of source positions, N, is chosen a priori to correspond to the total number of reflectors present in the environment. Secondly, the source is placed in a controlled position at each step, meaning that the first two peaks in $h_{i,m}$ always correspond to the direct-path and first-order reflection, with respect to each particular source position, microphone array and reflector arrangement. Evidently, $h_{i,m}$ contains information related to more than one reflector. Exploiting such redundancy is indeed possible, such as proposed by the authors in [114]. However, for the purposes of this manuscript we do not aim at exploiting redundancy or the reduction of the amount of source positions probed.

The 2-D reflector localization techniques outlined in [113–115] are extended to the 3-D case in a straightforward yet effective way. A 3-D microphone array, such as depicted in Figure 5.1(a), is employed to capture $h_{i,m}$. The 3-D space is decomposed into three orthogonal 2-D regions by considering three subsets of microphones. Let $\mathbf{r}_{\mathrm{xy}}$, $\mathbf{r}_{\mathrm{xz}}$, $\mathbf{r}_{\mathrm{yz}}$ denote the subsets lying on the xy, xz and yz-planes, as shown in Figures 5.1(b)–5.1(d) respectively. Each subset is used to identify the line parameters of the reflectors coincident with the plane it can observe. By combining the measurements from all three planes, each reflector plane is represented by a pair of lines lying on two orthogonal planes. By first estimating the line parameters of the reflectors it is possible to then calculate the parameters of the planes that are coincident with the actual reflectors in 3-D. In the following section we will outline the estimation of the line parameters of the reflectors followed by the reflector plane estimation methodology.

5.3.2 Disambiguation of Rectangular Patterns in the Hough Space

Consider the Hough parameter space representation of the estimated line reflectors. Let $H_1 = (\rho_1, \theta_1), H_2 = (\rho_2, \theta_2), \cdots, H_v = (\rho_v, \theta_v)$ denote the v peaks of $\mathcal{A}\left(\hat{\rho}, \hat{\theta}\right)$ [141]. Peaks H_m and H_n are paired together if they satisfy:

$$|\theta_m - \theta_n| \approx T_\theta, \tag{5.9}$$

where T_θ is an angular threshold, and determines if peaks H_m and H_n correspond to orthogonal lines (i.e. $T_\theta \approx \pi/2$).

For every microphone sub-array the peaks in the Hough accumulator are sorted with respect to (5.9). In this way the TOAs that are not related to a particular sub-array, and its respective plane, can be discarded. The disambiguation approach above is only valid for rectangular geometries. The interested reader is referred to [159] where other geometrical relationships are considered in order to perform disambiguation in more complex geometries.

5.3.3 Reflector Plane Estimation

The estimation process outlined in the previous Section leads to 6 pairs of reflector lines, one pair for each wall. In particular, each wall is represented by a pair of lines lying on two orthogonal planes. The reflector, therefore, has to be estimated as the plane which best fits the two lines.

We proceed as follows. With reference to Figure 5.2, let us consider two arbitrary lines $\mathbf{l}_1$ and $\mathbf{l}_2$. We aim at estimating the plane $\mathbf{P} = [\mathbf{n},\ d]^T$ that best fits $\mathbf{l}_1$ and $\mathbf{l}_2$, where the unit vector $\mathbf{n}$ is the normal of the plane, and d is its distance from the origin. For each line we select two arbitrary points lying on it, namely $\mathbf{p}_1$ and $\mathbf{q}_1$ on $\mathbf{l}_1$; and $\mathbf{p}_2$ and $\mathbf{q}_2$ on $\mathbf{l}_2$. We organize the point coordinates in the matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{p}_1 & 1 \\ \mathbf{q}_1 & 1 \\ \mathbf{p}_2 & 1 \\ \mathbf{q}_2 & 1 \end{bmatrix} . \tag{5.10}$$

The searched plane is then estimated, in the least-squares sense, as

$$\hat{\mathbf{P}} = [\hat{\mathbf{n}}, \hat{d}]^T = \arg\min_{\mathbf{n},d} \|\mathbf{G}[\mathbf{n},\ d]^T\|^2 \quad \text{s.t.} \quad \|\mathbf{n}\| = 1 \, . \tag{5.11}$$

The result of the above minimization yields the parameters of the plane that best describes the actual location of the reflector.

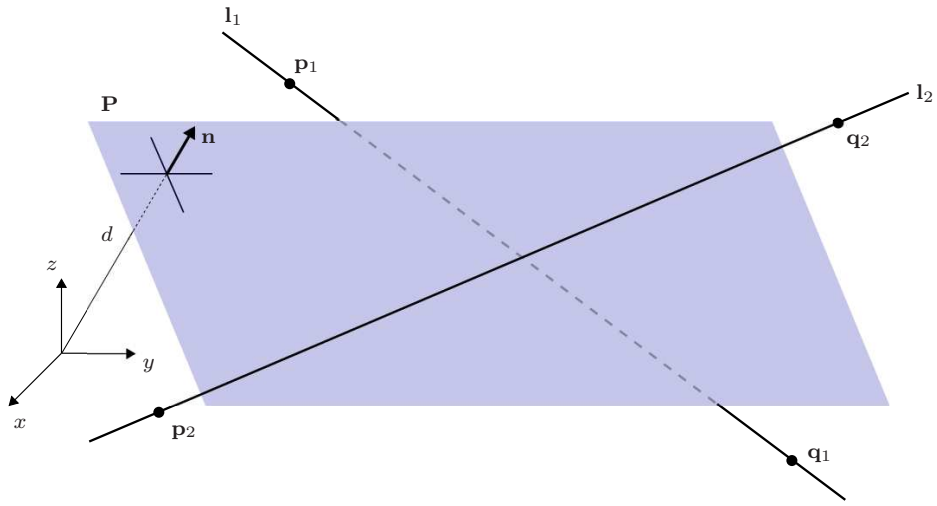


Figure 5.2: Plane estimation from two skew lines.

5.4 Direct Reflector Plane Localization

The following approach follows the derivation first presented in [117]. Reflectors are represented in 3-D by the coordinates of the plane on which they lie. Points on the plane $\mathbf{P} = [p_1\ p_2\ p_3\ p_4]^T$ satisfy the equation $p_1x + p_2y + p_3z + p_4 = \mathbf{P}^T\mathbf{x} = 0$, where $\mathbf{x} = [\lambda x\ \lambda y\ \lambda z\ \lambda]^T, \lambda \neq 0$, are the homogeneous coordinates for the point $[x\ y\ z]^T$ lying on the plane. Using the parameters $\{a, b, c, d, e, f, g, h, k, l\}$ a quadric can be expressed as [144]

$$\mathcal{Q} = \left\{(x, y, z) \in \mathbb{R}^3 | ax^2 + bxy + cy^2 + dxz + eyz + fz^2 + gx + hy + kz + l = 0\right\}. \tag{5.12}$$

The more compact matrix representation of (5.12) is given by

$$\mathbf{x}^T\mathbf{Q}\mathbf{x} = 0, \tag{5.13}$$

where the quadric matrix is defined as

$$\mathbf{Q} = \begin{bmatrix} a & b & d & g \\ b & c & e & h \\ d & e & f & k \\ g & h & k & l \end{bmatrix}.$$

The problem of localizing reflectors in 3D corresponds to finding the parameters $\mathbf{P}$ of the planes on which the reflectors lie, given estimates $\hat{h}_{i,m}(n)$. When multiple reflectors

are present in the acoustic scene, multiple independent estimations are performed, by sequentially placing the source in the proximity of each wall to be localized. We momentarily drop the index m, that describes each independent measurement, and focus only on the reflector planes of the ith channel. The aim now is to derive the equation for an ellipsoid that is tangential to the unknown reflector plane.

The implicit equation of an ellipsoid with foci in $(x_\mathrm{s}, y_\mathrm{s}, z_\mathrm{s})$ and (x_i, y_i, z_i) and with major diameter $r_i = \mathfrak{c}\tau_i^{\mathrm{RE}}$ is

$$\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}+\sqrt{(x-x_\mathrm{s})^2+(y-y_\mathrm{s})^2+(z-z_\mathrm{s})^2}=r_i\ . \quad (5.14)$$

By expanding (5.14) and comparing term by term with a quadric with parameters $(a_i, b_i, c_i, d_i, e_i, f_i, g_i, h_i, k_i, l_i)$ we obtain

$$\begin{aligned}
a_i &= 4\left[(x_\mathrm{s}-x_i)^2-r_i^2\right]\ ,\\
b_i &= 8\left[(x_\mathrm{s}-x_i)(y_\mathrm{s}-y_i)\right]\ ,\\
c_i &= 4\left[(y_\mathrm{s}-y_i)^2-r_i^2\right]\ ,\\
d_i &= 8\left[(x_\mathrm{s}-x_i)(z_\mathrm{s}-z_i)\right]\ ,\\
e_i &= 8\left[(y_\mathrm{s}-y_i)(z_\mathrm{s}-z_i)\right]\ ,\\
f_i &= 4\left[(z_\mathrm{s}-z_i)^2-r_i^2\right]\ ,\\
g_i &= 4\left[r_i^2(x_\mathrm{s}+x_i)-(x_i-x_\mathrm{s})(x_\mathrm{s}^2-x_i^2+y_\mathrm{s}^2+z_\mathrm{s}^2-y_i^2-z_i^2)\right]\ ,\\
h_i &= 4\left[r_i^2(y_\mathrm{s}+y_i)-(x_i-x_\mathrm{s})(x_\mathrm{s}^2-x_i^2+x_\mathrm{s}^2+z_\mathrm{s}^2-x_i^2-z_i^2)\right]\ ,\\
k_i &= 4\left[r_i^2(z_\mathrm{s}+z_i)-(x_i-x_\mathrm{s})(z_\mathrm{s}^2-z_i^2+y_\mathrm{s}^2+x_\mathrm{s}^2-x_i^2-y_i^2)\right]\ ,\\
l_i &= \left[(x_\mathrm{s}^2+y_\mathrm{s}^2+z_\mathrm{s}^2)+(x_i^2+y_i^2+z_i^2-r_i^2)\right]^2-4(x_\mathrm{s}^2+y_\mathrm{s}^2+z_\mathrm{s}^2)(x_i^2+y_i^2+z_i^2)\ .
\end{aligned}$$

More meaningful for our purposes is the definition of the *plane conic*. The plane $\mathbf{P}$ is tangential to the quadric $\mathbf{Q}_i$ iff

$$\mathbf{P}^T\mathbf{Q}_i^*\mathbf{P}=0, \quad (5.15)$$

where $\mathbf{Q}_i^* = \det(\mathbf{Q}_i)\mathbf{Q}_i^{-1}$ is the adjoint of the quadric matrix $\mathbf{Q}_i$.

5.4.1 Constrained Least-Squares Solution

Combining the constraints in (5.15), the reflector plane is the simultaneous solution of [113, 114]

$$\begin{cases} \mathbf{P}^T\mathbf{Q}_0^*\mathbf{P} & = 0 \\ \mathbf{P}^T\mathbf{Q}_1^*\mathbf{P} & = 0 \\ \cdots & \\ \mathbf{P}^T\mathbf{Q}_{M-1}^*\mathbf{P} & = 0 \end{cases} . \tag{5.16}$$

Since we have four unknowns (the parameters p_1, p_2, p_3, p_4) we need at least $M = 4$. The cost function that combines the individual constraints is given by [117]

$$J\left(\mathbf{l}, \{\mathbf{Q}_i^*\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} \left\|\mathbf{P}^T\ \mathbf{Q}_i^*\ \mathbf{P}\right\|^2,$$

and the four unknown plane parameters can be estimated by minimizing the cost function

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} J\left(\mathbf{P}, \{\mathbf{Q}_i^*\}_{i=0}^{M-1}\right) \quad \text{s.t.} \quad \|\mathbf{P}\| = 1\ . \tag{5.17}$$

Note, that in order to avoid the trivial solution $\mathbf{P} = \mathbf{0}$, the unitary norm is imposed on the solution.

The above estimator is equivalent to the constrained least-squares solution outlined in section 4.2 and adopted in [113, 114].

5.4.2 Exact solution

The cost function, outlined in the previous section, is a fourth-order polynomial. It will be shown that it is possible to reformulate it as a second-order polynomial with a single quadratic constraint, that admits an exact solution. This approach guarantees that the true reflector is found and local minima are avoided.

Without loss of generality, we now translate the reference frame at the source, so that its position becomes $\mathbf{r}_\mathrm{s} = [0\ \ 0\ \ 0]^T$. This simple transformation will turn out useful in determining an exact solution to the reflector localization problem. We notice, however, that the resulting estimate $\hat{\mathbf{P}}$ of the reflective plane refers to the source position, which changes for the different walls to be localized. Afterwards, therefore, we convert the estimated plane vectors to the original coordinate system.

It is convenient to adopt the dual representation of the quadric, namely $\mathbf{P}^T\mathbf{Q}_i^*\mathbf{P} = 0$, which is satisfied by all the planes $\mathbf{P} = [p_1\ p_2\ p_3\ p_4]^T$ of equation $p_1x+p_2y+p_3z+p_4 = 0$

tangential to the ellipsoid. Computing $\mathbf{Q}_i^* = \det(\mathbf{Q}_i)\mathbf{Q}_i^{-1}$ we obtain

$$\mathbf{Q}_i^* = \begin{bmatrix} \alpha_i^* & 0 & 0 & \frac{\delta_i^*}{2} \\ 0 & \alpha_i^* & 0 & \frac{\eta_i^*}{2} \\ 0 & 0 & \alpha_i^* & \frac{\iota_i^*}{2} \\ \frac{\delta_i^*}{2} & \frac{\eta_i^*}{2} & \frac{\iota_i^*}{2} & \kappa_i^* \end{bmatrix}, \quad \begin{aligned} \alpha_i^* &= 16r_i^4(\|\mathbf{r}_i\|^2 - r_i^2)^2\,, \\ \delta_i^* &= 64r_i^4 x_i(\|\mathbf{r}_i\|^2 - r_i^2)\,, \\ \eta_i^* &= 64r_i^4 y_i(\|\mathbf{r}_i\|^2 - r_i^2)\,, \\ \iota_i^* &= 64r_i^4 z_i(\|\mathbf{r}_i\|^2 - r_i^2)\,, \\ \kappa_i^* &= 64r_i^4(\|\mathbf{r}_i\|^2 - r_i^2)\,. \end{aligned}$$

Combining the TOA measurements leads to the definition of the minimization problem

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} J(\mathbf{P}) = \arg\min_{\mathbf{P}} \sum_{i=0}^{M-1} \left\|\mathbf{P}^T\, \mathbf{Q}_i^*\, \mathbf{P}\right\|^2, \tag{5.18}$$

where the plane of reflection is estimated as the global minimum of the cost function $J(\mathbf{P})$, which is the sum of the squared residuals of all the quadratic constraints. Following the same approach proposed in [116] for reflector line estimation, we restrict the search space to planes having $p_4 = 1$. This means discarding all the planes passing through the origin, which can not generate any reflective path since they contain the source. As a result, we obtain

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} \sum_{i=0}^{M-1} \left[\alpha_i^*(p_1^2 + p_2^2 + p_3^2) + \delta_i^* p_1 + \eta_i^* p_2 + \iota_i^* p_3 + \kappa_i^*\right]^2 . \tag{5.19}$$

By posing $w = p_1^2 + p_2^2 + p_3^2$, (5.19) can be rewritten as [152]

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left\{\|\mathbf{A}\mathbf{w} - \mathbf{b}\|^2 : \mathbf{w}^T\mathbf{D}\mathbf{w} + 2\mathbf{f}^T\mathbf{w} = 0\right\} , \tag{5.20}$$

where $\mathbf{w} = [w\ p_1\ p_2\ p_3]^T$ and

$$\mathbf{A} = \begin{bmatrix} \alpha_0^* & \delta_0^* & \eta_0^* & \iota_0^* \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{M-1}^* & \delta_{M-1}^* & \eta_{M-1}^* & \iota_{M-1}^* \end{bmatrix}, \quad \begin{aligned} \mathbf{b} &= -[\kappa_0^* \ldots \kappa_{M-1}^*]^T\,, \\ \mathbf{D} &= \mathrm{diag}(0,1,1,1)\,, \\ \mathbf{f} &= [-0.5\ 0\ 0\ 0]^T\,. \end{aligned}$$

The exact solution $\hat{\mathbf{w}} = [\hat{w}\ \hat{p}_1\ \hat{p}_2\ \hat{p}_3]^T$ to (5.20) can be found efficiently as in [116, 152], and the searched plane of reflection is given by $\hat{\mathbf{P}} = [\hat{p}_1\ \hat{p}_2\ \hat{p}_3\ 1]^T$, which is expressed in the source reference system. The solution in the original reference system is finally obtained as $\hat{\bar{\mathbf{P}}} = (\mathbf{T}_\mathrm{s}^{-1})^T\hat{\mathbf{P}}$, where

$$\mathbf{T}_\mathrm{s} = \begin{bmatrix} 1 & 0 & 0 & x_\mathrm{s} \\ 0 & 1 & 0 & y_\mathrm{s} \\ 0 & 0 & 1 & z_\mathrm{s} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

denotes the translation from the original reference frame to that centered at $\mathbf{r}_\mathrm{s} = [x_\mathrm{s}\ y_\mathrm{s}\ z_\mathrm{s}]^T$.

5.4.3 Theoretical Error Analysis

The theoretical analysis has been performed using the setup shown in Figure 5.3. For the sake of clarity in the analysis of the results, in this section we parametrize the reflective plane by its distance d from the origin; the azimuth ϕ; and the co-elevation θ. The

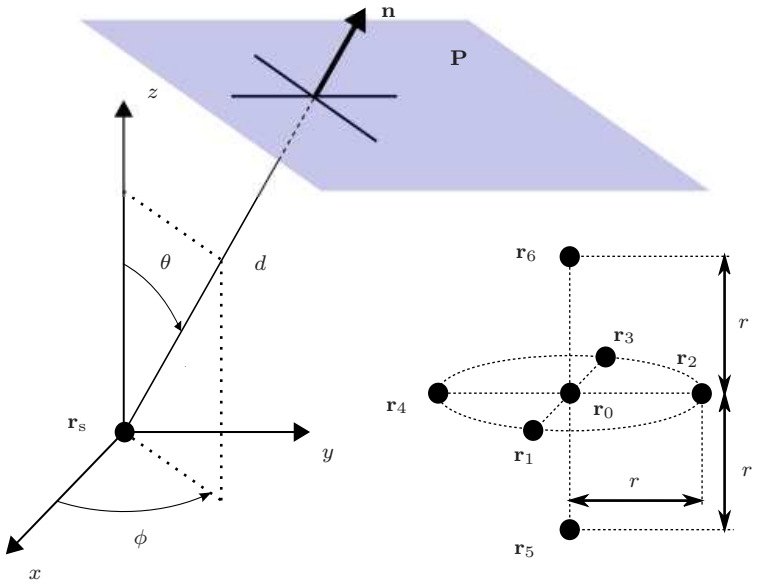


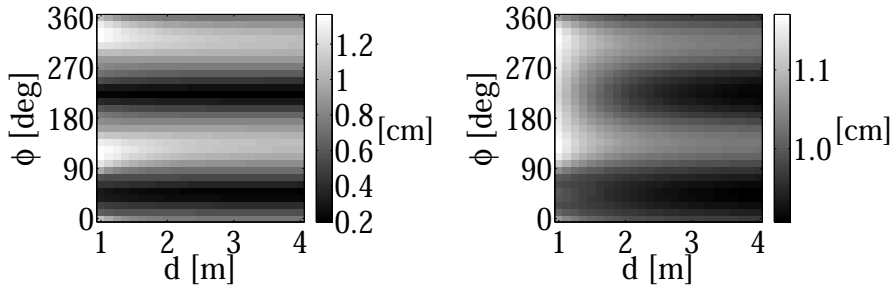
Figure 5.3: Evaluation setup

plane parametrization adopted in section 5.4.2 is related with the current parametrization through $\mathbf{P} = [\mathbf{n}^T, d]^T$, $\mathbf{n} = [\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta]^T$ being the unit vector normal to the plane. The microphone array accommodates 7 sensors deployed as in Figure 5.3. In particular, the central microphone is located at $\mathbf{r}_0 = [0.5\ 0.5\ 0]^T$, and the remaining sensors are all located at a distance $r = 0.25\,\text{m}$ from $\mathbf{r}_0$. If $\hat{d}$, $\hat{\phi}$ and $\hat{\theta}$ are the estimated plane parameters, then the localization accuracy is assessed in terms of the distance error $\epsilon_\text{d} = \left|d - \hat{d}\right|$; the azimuth error $\epsilon_\phi = \left|\phi - \hat{\phi}\right|$; and the co-elevation error $\epsilon_\theta = \left|\theta - \hat{\theta}\right|$.

The analysis has been carried out for a set of 1500 test reflector positions, whose parameters vary on a multidimensional grid defined by: 25 values for the distance d in the range $[1\,\text{m} \sim 4\,\text{m}]$; 30 values for the azimuth ϕ the range $[0 \sim 2\pi]$; and 2 points for the co-elevation, namely $\theta = \pi/2$ and $\theta = \pi/6$. We assumed the error on TOA measurements to be zero-mean and Gaussian distributed with standard deviation

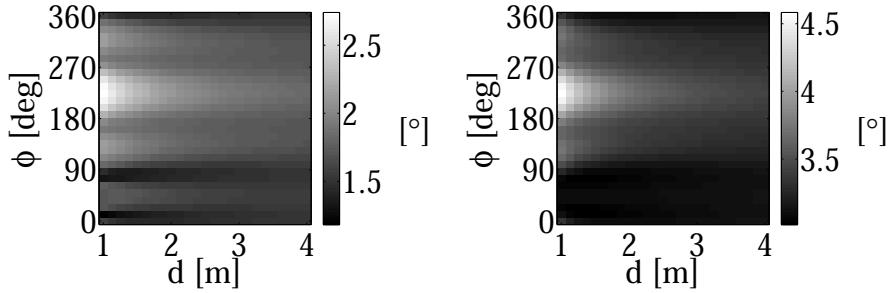
$\sigma = 0.01\,\mathrm{m}/c$, independent on each microphone.

Figure 5.4 shows the resulting standard deviation of the distance error, for all the tested reflector planes. In particular, Figure 5.4-(a) is relative to the case of $\theta = \pi/2$; and Figure 5.4-(b) to $\theta = \pi/6$. Similarly, Figures 5.5–5.6 depict the theoretical standard deviation of the azimuth error and the co-elevation error, respectively. Interestingly,



(a) $\theta = \pi/2$ (b) $\theta = \pi/6$

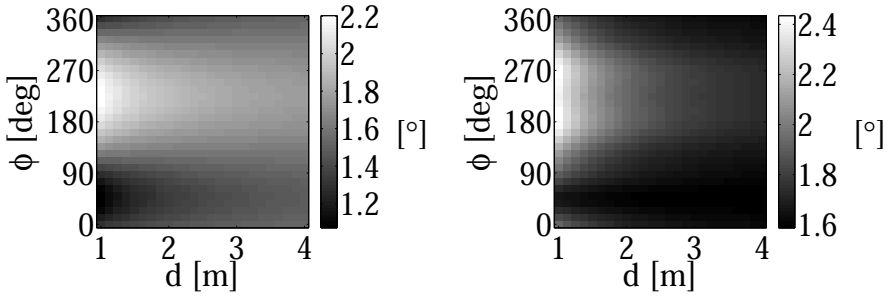
Figure 5.4: Theoretical standard deviation of ϵ_{d}.



(a) $\theta = \pi/2$ (b) $\theta = \pi/6$

Figure 5.5: Theoretical standard deviation of ϵ_{ϕ}.

for the setup under analysis, the localization accuracy is almost independent from the distance d of the reflector, while being highly variable with the azimuth ϕ. In particular, from Figure 5.4 we observe that, for planes having $\phi \in [0°, 90°]$ or $\phi \in [180°, 270°]$, ϵ_{d} tends to be low while being higher for the other azimuth angles. Conversely, we notice from Fig. 5.5 that ϵ_{ϕ} exhibits an opposite behaviour, being higher when $\phi \in [0°, 90°]$ or $\phi \in [180°, 270°]$. Figure 5.6 reveals that ϵ_{θ} has a smoother behaviour, but still presents higher values for planes with $\phi \in [180°, 270°]$. Finally, we observe that the accuracy



(a) $\theta = \pi/2$ (b) $\theta = \pi/6$

Figure 5.6: Theoretical standard deviation of ϵ_θ.

tends to decrease for elevated planes, especially for the azimuth error ϵ_ϕ, whose average is 1.74° for $\theta = \pi/2$ and 3.34° for $\theta = \pi/6$.

5.5 Experimental Verification

The two planar reflector localization algorithms were verified by two experiments in a real shoebox-shaped conference room measuring $L \times W \times H = 2.77\text{m} \times 3.55\text{m} \times 3.17\text{m}$, built with concrete walls and ceiling and floored with linoleum floor covering.

5.5.1 Evaluation Criteria

Regarding Experiment 1, the accuracy of the reflector localization is assessed in two steps. First the estimated line parameters in each of the three planes are compared to the hand measured ground truths in terms of a distance and angular error. As a next step the estimated plane parameters are compared to the true planes in terms of a point-plane distance and their dihedral angle.

Let $\mathbf{l}$ and $\hat{\mathbf{l}}$ be the reflector line and its estimate, respectively. From these we can evaluate the distance d from $\mathbf{r}_0$ to each line and the orientation α. The distance and orientation can be evaluated by projecting $\mathbf{r}_0$ onto the line such that $d = \frac{|l_1\, x_0 + l_2\, y_0 + l_3|}{\sqrt{l_1^2 + l_2^2}}$, $\alpha = \arctan \frac{l_2}{l_1}$. The accuracy of the reflector localization is measured in terms of distance error $\epsilon_\text{d} = \left| d - \hat{d} \right|$ and angular error $\epsilon_\text{a} = |\alpha - \hat{\alpha}|$. The distance and angular error results for the reflector lines in 2-D are shown for the xy, xz and yz planes in table 5.1.

Let $\mathbf{P}$ and $\hat{\mathbf{P}}$ be the true and estimated reflector planes respectively. From these we can evaluate the distance $d^{3\mathrm{D}}$ from $\mathbf{r}_0$ to each plane (point-plane distance) and the angle between true and estimate planes (dihedral angle). The distance is given by $d^{3\mathrm{D}} = \mathbf{n}^T[x_0\ y_0\ z_0] + d'$, where $\mathbf{n}$ is the unit normal vector of $\mathbf{P}$, and d' is the constant of the Hessian normal form. The distance error is calculated as $\epsilon_{\mathrm{D}} = \left|d^{3\mathrm{D}} - \hat{d^{3\mathrm{D}}}\right|$ and the dihedral angle is given by $\Phi = \arccos(\mathbf{n}^T\hat{\mathbf{n}})$.

Regarding Experiment 2, we assume that $\hat{d}$, $\hat{\phi}$ and $\hat{\theta}$ are the estimated plane parameters, then the localization accuracy is assessed in terms of the distance error $\epsilon_{\mathrm{d}} = \left|d - \hat{d}\right|$; the azimuth error $\epsilon_\phi = \left|\phi - \hat{\phi}\right|$; and the co-elevation error $\epsilon_\theta = \left|\theta - \hat{\theta}\right|$.

Experiment 1

The central microphone of the array is placed at a distance of 1.2m from the West wall, 1.91m from the South wall and 1.59m from the floor. The array is composed of 7 omnidirectional microphones. On the horizontal plane the extension of the array is 0.5m $\times$ 0.5m, while the microphones on the vertical axis are kept 0.38cm apart. The room impulse response was measured using the MLS method [43] from 6 different locations using a sampling rate of 48 kHz. The sequence is then processed to extract the impulse response from each position of the source to each microphone in the array.

The results for the plane localization in 2-D and 3-D are shown in Tables 5.1–5.2. The ground-truth distances were measured to within an estimated measurement error. Nonetheless, the localization for both the line reflectors and the planar reflectors is to within a few cm accurate. It is important to note however that errors propagate directly from the estimation of the linear reflectors to the localization of the planar reflectors.

Experiment 2

The microphone array used in the experiment has the same geometry of that in Figure 5.3, and the central microphone was placed at a distance of 1.2 m from the West wall; 1.91 m from the South wall; and 1.59 m from the floor. The room impulse response was measured using the MLS method [43] from 6 different locations using a sampling rate of 48 kHz. Consequently, $N = 7$ impulse responses are measured for each wall.

The experimental results are shown in Table 4.4, giving on a per wall basis, the distance error ϵ_{d}; the azimuth error ϵ_ϕ; and the co-elevation error ϵ_θ. The results are

Table 5.1: Experiment 1: Distance and angular error results in each plane

	xy		xz		yz	
Reflector	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]
West	1.180	0.619	2.720	0.871		
South	1.030	1.598			4.810	0.034
East	1.310	0.160	7.300	0.504		
North	1.690	0.160			3.100	0.756
Ceiling			1.620	0.728	2.780	0.275
Floor			1.030	1.598	0.210	0.092

Table 5.2: Experiment 1: Distance and angular error comparison in 3-D

Reflector	ϵ_D [cm]	Φ [°]
West	0.260	1.068
South	1.300	0.718
East	7.950	0.528
North	0.063	0.773
Ceiling	1.050	1.601
Floor	3.860	0.778

expressed with respect to the central microphone position $\mathbf{r}_0$. All the reflectors are localized to within a few centimeters. On average, the distance error is 1.5 cm, the azimuth error 0.87°, and the co-elevation error is 0.76°. It is important to note that, for reflectors having $\theta = 0°$ and $\theta = 180°$ (ceiling and floor), the azimuth is not defined and therefore the co-elevation error fully characterizes the angular accuracy.

Conclusion

This chapter extended the 2-D reflector localization approach introduced in Chapter 4 to the 3-D case. First, we presented an approach for estimating the geometry of an acoustic enclosure by transforming the localization of planar reflectors into the estimation of multiple linear reflectors. Experimental results in a real conference room

Table 5.3: Experiment 2: Distance, azimuth and co-elevation error

Reflector	$(d\,,\phi\,,\theta)$	ϵ_{d} [cm]	ϵ_{ϕ} [°]	ϵ_{θ} [°]
West	$(1.20\,\mathrm{m}\,,180^{\circ}\,,90^{\circ})$	1.3	0.49	0.62
South	$(1.91\,\mathrm{m}\,,270^{\circ}\,,90^{\circ})$	0.1	1.81	1.89
East	$(1.57\,\mathrm{m}\,,0^{\circ}\,,90^{\circ})$	1.1	1.09	0.22
North	$(1.64\,\mathrm{m}\,,90^{\circ}\,,90^{\circ})$	4.3	0.10	0.71
Ceiling	$(1.58\,\mathrm{m}\,,-\,,0^{\circ})$	0.9	–	0.72
Floor	$(1.59\,\mathrm{m}\,,-\,,180^{\circ})$	1.3	–	0.38

demonstrate the feasibility of the proposed method. As a second step, the problem was solved entirely in 3-D using a closed-form estimator. The analysis of the data acquired in a small conference room confirms that the proposed technique is able to localize the reflectors with a distance error of a few centimeters and an angular error below one degree.

6

Applications and Future Work

Introduction

INFERRING information about the environment conditions in which space-time processing algorithms operate is an emerging research topic. This information can be useful for source identification [160], source localization [80], speech enhancement [161], dereverberation [2], echo cancellation [162] and wave field rendering [61, 163, 164].

In summary, up to this point in the manuscript the required prior information needed for the reflector localization method included: i) the relative array geometry, ii) an estimate of the speed of sound, iii) a representation of the acoustic impulse responses (AIRs) and related parameters (such as the sampling frequency). In practice, the AIR capturing process is known and the sensors are placed at predefined, but arbitrary, positions. The remaining missing information that needs to be inferred as a prerequisite for robust time-of-arrival (TOA) estimation is the speed of sound.

The quintessential goal of geometric room inference, and environment parameter estimation in general, is to infer as much as possible about the acoustic environment from as little information as possible (i.e. minimum prior assumptions). It is evident that in the methods outlined so far, there exists a natural order of dependencies: acoustic reflector localization methodologies rely on accurate time-difference-of-arrival (TDOA) or TOA information that in turn dependent on accurate sound source localization and consequently on an accurate estimate of the actual speed of sound. This hierarchical top-down order can be reformulated from the bottom-up to state that accurate speed

of sound estimation leads to accurate sound source position estimation which in turns leads to accurate reflector localization. The estimation of the actual speed of sound is consequently the vital first step on which all other procedures rely.

In this chapter, first a methodology for blindly estimating the speed of sound using only TDOA measurements is outlined. Accurate estimates of the actual sound speed are desired since environmental properties, such as the ambient temperature, directly influence the propagation speed of the sound waves. Most acoustic processing algorithms assume a known propagation speed, which is a reliable assumption only under controlled laboratory conditions. Accurately estimating the speed of sound is therefore highly desired. In uncontrolled environments, the use of the standard speed value might lead to inaccuracy in the source localization due to temperature variations [165].

As a next step in this chapter, a discussion on the geometric inference algorithm is presented that includes suggestions for future work. As outlined so far, the reflector localization yields the position of the acoustic reflectors in the environment, however it does not yield other important properties and characteristics of the acoustic reflectors. First and foremost, the absorption on the reflectors and associated signal decay rates are not considered. Additionally, scattering effects are not taken into account. A clear discussion and plan of action regarding these issues is presented with possibilities for future work.

The chapter is organized as follows: Section 6.1 introduces a geometric inference methodology robust to variations in temperature and consequently TOA and TDOA information. Section 6.2 discusses possible extensions to the geometric inference process.

Relevant Publications

1. P. Annibale, J. Filos, P. A. Naylor and R. Rabenstein, "Geometric Inference of the Room Geometry Under Temperature Variations," *5th International Symposium on Communications, Control, and Signal Processing (ISCCSP)*, Rome, Italy, May 2 - 4, 2012.

2. P. Annibale, J. Filos, P. A. Naylor and R. Rabenstein, "TDOA-based Speed Of Sound Estimation for Air Temperature and Geometric Inference," *IEEE Trans. Audio, Speech, Lang. Process.*, 2013.

6.1 Geometric Inference under Temperature Variations

The dependence of the speed of sound on the temperature of the propagation medium is well known in acoustics. This dependence on average temperature and spatial temperature distributions has been investigated in many application areas. A few examples are fault diagnosis of a liquid transmission system [166], ultrasound diagnosis [167], wind velocity and speed of sound estimation with an acoustic ranging system [168], acoustic propagation in oceanic environments [169], and the measurement of air temperature and wind velocity by acoustic tomography [170]. In this section the geometric inference framework is combined with a method that blindly and robustly estimates the velocity of the speed of sound [171, 172].

Under temperature variations a standard value for the speed of sound might yield an inaccurate estimation of the TOAs $\tau_{i,k}$ impairing the reflector localization, in this case an estimate of the actual speed of sound is necessary. A novel method to estimate the speed of sound was presented in [173]. Such a method relies merely on TDOA measurements and therefore it is suited for geometric inference when no synchronization between source and receiver is available.

6.1.1 Speed of Sound Estimation from TDOAs

Consider the two-dimensional inference problem, i.e. an acoustic source lies in an unknown position $\mathbf{r}_\mathrm{s}$ and M sensors are distributed at the known positions $\mathbf{r}_i$ with $i = 0, \cdots, N$ and $N = M - 1$. From the M estimated AIRs a *spherical* set of N TDOAs can be obtained as time differences between the direct-path peaks and the direct-path peak of the reference microphone. If the first microphone $\mathbf{r}_0$ is chosen as reference such a set may be represented by the following N-vector

$$\boldsymbol{\tau} = \frac{1}{f_s} \begin{bmatrix} n_{1,0} - n_{0,0} \\ n_{2,0} - n_{0,0} \\ \cdots \\ n_{N,0} - n_{0,0} \end{bmatrix} . \tag{6.1}$$

According to [173] the following scalar function of the assumed signal propagation speed c can be written

$$\delta(c) = ||\boldsymbol{\Gamma} \boldsymbol{b}(c)|| - \frac{1}{c} \boldsymbol{\Theta} \boldsymbol{b}(c) , \tag{6.2}$$

where the constant matrices $\mathbf{\Theta}$, $\mathbf{\Gamma}$ and the N-vector $\boldsymbol{b}(c)$ depend only on the microphone positions $\mathbf{r}_i$, $, i = 1, \cdots, N$ and the vector $\boldsymbol{\tau}$. The zero of the above function is an estimate of the actual propagation speed, in this case the actual speed of sound.

Unfortunately such a function $\delta(c)$ involving the Euclidean norm of $\mathbf{\Gamma}\boldsymbol{b}(c)$ is non-linear, therefore applying a root-finding algorithm might be intractable. This issue can be overcome by linearizing $\delta(c)$ near to its zero-crossing as it has been shown to be approximately linear in that range. In acoustic applications, the standard value of speed of sound at 20 °C may be chosen as a reliable linearization point $\bar{c}$, leading to the following first order Taylor expansion

$$\delta(c) \approx \delta_{\text{lin}}(c) = \bar{\delta} + \bar{\delta}'(c - \bar{c}) \,, \tag{6.3}$$

with

$$\bar{\delta} = \delta(\bar{c}) \quad \text{and} \quad \bar{\delta}' = \left.\frac{\mathrm{d}\delta(c)}{\mathrm{d}c}\right|_{c=\bar{c}} . \tag{6.4}$$

Finally, the estimated propagation speed value is given by the zero-crossing point $\hat{c}$ of the linearized function $\delta_{\text{lin}}(c)$, i.e.

$$\hat{c} = \frac{\bar{\delta} - \bar{\delta}'\bar{c}}{\bar{\delta}'}, \tag{6.5}$$

where the value of the first order derivative $\bar{\delta}'$ at $\bar{c}$ can be calculated with derivation rules from (6.2).

6.1.2 Multiple Sources Approach

As reported in [173], the above speed estimate can be improved in noisy conditions by exploiting the *full* TDOA set. However here the TDOAs are not obtained by signal correlation, rather they are extracted from the M estimated AIRs. As a consequence the construction of the *full* TDOA set will not add any useful information for the speed estimation. Moreover the microphone array used for the geometric inference experiments (see Section 6.1.3) uses only $M = 5$ microphones, i.e. only one microphone provides redundant information to counteract the effect of the measurement noise since the minimum number of sensors required for the two dimensional speed estimation problem is 4. Nevertheless the robustness of the algorithm can be still improved by assuming that for a small-size array the speed of sound in a reasonable time interval is the same regardless of the source position. Following this idea the scalar function

$\delta(c)$ in (6.2) can be built using TDOA sets generated by different acoustic sources. For a rectangular room, four differently located acoustic sources are used, as shown in Figure 6.1). Hence for each source $\mathbf{r}_j$ $(j = 1, 2, 3, 4)$ a function $\delta_j(c)$ is derived. In noisy conditions a robust speed of sound estimate can be found form the minimization in the least-squares sense of such functions. The corresponding cost function is given by

$$\sum_{j=1}^{4} \delta_j(c)^2 = \sum_{j=1}^{4} \left(||\mathbf{\Gamma}_j \boldsymbol{b}_j(c)|| - \frac{1}{c} \mathbf{\Theta}_j \boldsymbol{b}_j(c) \right)^2 , \tag{6.6}$$

where the index j indicates that the matrices $\mathbf{\Theta}$, $\mathbf{\Gamma}$ and the vector $\boldsymbol{b}(c)$ are obtained using the vector $\boldsymbol{\tau}_j$ corresponding to the source $\mathbf{r}_j$. Again the linear approximation described in Section 6.1.1 can be applied to perform the minimization efficiently.

The resulting speed of sound estimate can be now used to accurately estimate the TOAs $\tau_{i,k}$, by means of source localization algorithms [121], and then perform the reflector localization.

6.1.3 Experimental Results

The effects of temperature variation on the speed of sound within the reflector localization framework have been evaluated in a real conference room measuring 3.31 × 3.58 × 3.00 m, with concrete walls and two flush-mounted wooden doors in the south and east walls. A microphone array consisting of four microphones spaced by 0.5 m in a '+' configuration and a fifth placed in the centre was positioned at (1.75,1.5) m from the south-west corner. A Genelec 8030A loudspeaker was placed at four distinct source positions. The microphone signals were sampled at 96 kHz. At each position, the acoustic impulse response between the source and microphone array was estimated using the MLS method [140].

At first, geometric inference has been performed without knowledge of the true speed of sound. As usual a value for the speed of sound has to be assumed to convert estimated time differences into range differences. Here a value of $c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$ has been adopted corresponding to a temperature of $\vartheta = 72\,^{\circ}\mathrm{C}$. This rather high value has been chosen to demonstrate the effect of an erroneous assumption.

Next the same set of TDOAs has been used for geometric inference, but this time the speed of sound at the time of measurement has been inferred from these TDOAs. The resulting speed of sound was $\hat{c} = 345\ \frac{\mathrm{m}}{\mathrm{s}}$ corresponding to a temperature of $\vartheta = 23\,^{\circ}\mathrm{C}$.

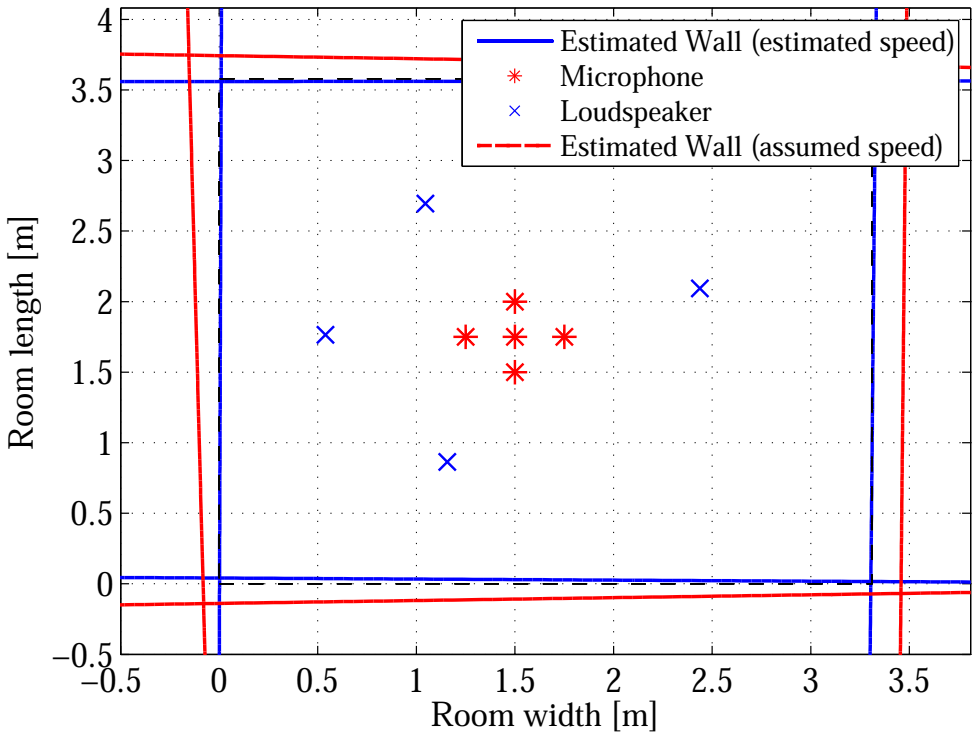


Figure 6.1: Room inference results using a microphone array, placed centrally in a small conference room, capturing a MLS sequence from 4 source positions in turn. Red lines: assumed speed of sound $c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$. Blue lines: estimated speed of sound $c = 345\ \frac{\mathrm{m}}{\mathrm{s}}$. Dashed black rectangle: actual geometry of the room.

The results in Figure 6.1 and Table 6.1 show not only the increase of the distance error ϵ_d and the angular error ϵ_a with an erroneously assumed temperature resp. speed of sound. Figure 6.1 displays also clearly the effect of a temperature increase on the inference of the wall positions: A higher value of the speed of sound virtually increases the size of the room by upscaling all involved distances.

Note that the room available for the experiment in Figure 6.1 is rather small. Larger rooms with a larger travel time of the reflections are more sensitive to speed of sound variations and exhibit the same absolute errors at smaller temperature variations.

A technique for inferring the geometry of a room under temperature variations has been presented. Since existing methods rely on either synchronized measurements or often inaccurate estimates of the speed of sound, the propagation speed is estimated

Table 6.1: Reflector localization results with real-world data

$c = 345\ \frac{\mathrm{m}}{\mathrm{s}}$			$c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$		
Wall	ϵ_d [cm]	ϵ_a [°]	Wall	ϵ_d [cm]	ϵ_a [°]
North	1.890	0.056	North	19.440	1.253
East	0.770	0.388	East	16.150	0.400
South	5.220	0.429	South	16.910	1.168
West	0.650	0.138	West	11.980	1.101

before the dominant reflectors are localized. In this way the room geometry can be reconstructed even when there are fluctuations in the ambient temperature. Improvements in accuracy are demonstrated in a real conference room that is exposed to strong variations in temperature.

6.2 Future Work

The methods established in this thesis consider the geometric localization of reflectors. Localizing the position of reflectors in turn yields the geometry, i.e. boundaries, of the acoustic enclosure. One might ask the question if the information about the geometry of the acoustic environment is enough to fully describe the propagation of sound in such an enclosed space. The short answer is no, but perhaps surprisingly there is more knowledge to be gained from the geometry of the acoustic scene than what initially meets the eye.

As mentioned in the beginning of this chapter, the propagation of sound, originating from a sound source inside an acoustic enclosure, depends primarily on the shape of the room. Other dependencies include the uniform or nonuniform absorption and scattering properties of the reflectors along with temperature, pressure and humidity of the propagating medium, amongst others. One of the most prominent acoustic characteristics of an enclosure is the reverberation time [2]. Its value can be used to predict speech intelligibility for example, and is used by speech enhancement techniques to suppress reverberation [161].

Determining absorption characteristics and sound source pressures in an enclosure have been extensively studied in the scientific literature [160, 174]. The prediction of

energy decay in room impulse responses has also been approached recently [175]. These methods rely on either beam tracing methods or an image-source model to approximate the energy-time curve. Given the geometry of the acoustic scene, the actual measured room impulse responses (RIRs), and a simulated representation of the RIRs, obtained by for example the image-method, our claim is that absorptions coefficients of surfaces can be inferred. Consider matching the actual energy decay curve (that is observed when a noise source is switched off) with the simulated energy decay curve of the RIR [54], then one might employ an iterative optimization procedure that estimates appropriate absorption coefficients of the reflectors. A similar methodology for blindly estimating the reverberation time based on the distribution of signal decay rates has already been proposed [161]. In summary, by inferring the geometry of the acoustic enclosure, the energy decay in impulse responses can be simulated from which the reverberation time and absorption characteristics can be inferred in turn.

Another important aspect that is only briefly mentioned in this manuscript involves the disambiguation of TOA and TDOA estimates from AIRs. The permutation and non-uniqueness problems in the TOA estimation have already been outlined in Section 4.3. Recently, a disambiguation method based on graph theory [176] has been proposed [154]. A possible extension of this work would be to employ the method in [154] to perform disambiguation of TOA and TDOA estimates. This could be used in conjunction with the disambiguation based on the Hough transform outlined in this work.

Conclusion

In this chapter, first a technique for inferring the geometry of a room under temperature variations was presented. Since existing methods rely on either synchronized measurements or often inaccurate estimates of the speed of sound, the propagation speed is estimated before the dominant reflectors are localized. In this way the room geometry can be reconstructed even when there are fluctuations in the ambient temperature. Experimental results in a real conference room, that was exposed to strong variations in temperature, demonstrated improvements in accuracy by blindly inferring the actual speed of sound first. As a next step, possible extensions of the geometric inference framework were considered. It was shown that additional acoustic characteristics, such

as the reverberation time or energy decay curves, could be inferred by taking into account the knowledge on the position of the reflectors in the acoustic environment.

7

Conclusions

IN this work, a novel methodology for inferring the geometry of an acoustic enclosure was developed and analyzed. The knowledge inferred about the acoustic characteristics of the environment can be advantageous for applications such as sound source localization, speech enhancement, dereverberation and adaptive echo cancellation by assisting in tracking environment changes and helping the initialization of such algorithms.

Having reviewed a variety of array processing and acoustic parameter estimation algorithms in Chapter 1, the concepts of system identification, room acoustics (i.e. simulation and measurement of room impulse responses), sound source localization and line estimation were introduced in Chapter 2.

As a first step, a geometric constraint [113, 114], that permits the estimation of the line parameters of reflectors in an acoustic scene, was introduced in Chapter 3. The geometric constraint was developed in a two-dimensional (2-D) plane, for illustration purposes. In this formulation time-of-arrival (TOA) and time-difference-of-arrival (TDOA) information, along with source and microphone positions, formed a set of elliptical constraints on the possible locations of the reflectors. It was shown that the common tangent of these constraints corresponds to the reflector location that was found by minimizing a specific cost function in the least-squares sense.

The localization of reflectors, such as the walls enclosing a room, in a 2-D plane, was presented in detail in Chapter 4. Specifically, the localization of multiple reflectors was achieved through estimation of the TOAs of reflected signals by analysis of acoustic impulse responses (AIRs) [3]. When multiple walls are present in the acoustic scene,

an ambiguity problem arises, which was addressed using the Hough transform [115]. Monte Carlo simulations revealed that the proposed method worked reliably even when the source location was not known and the AIRs were unsynchronized. Further simulations showed that by using the Hough transform, and taking sequential measurements for different source positions, the robustness to noise in the TOA information could be improved. Additionally, the performance of the estimators used in the reflector localization were compared theoretically and experimentally. Real-world measurements showed that the proposed technique provides reliable results in a practical setting involving plane wall reflectors forming the room.

The localization of reflectors was then extended to the three-dimensional (3-D) case in Chapter 5. Two methods were considered. First, a method that estimated planar reflector parameters, by combining multiple orthogonal line parameters, was presented [118]. More specifically, a 3-D array accommodating seven microphones was used. Microphones were organized in three sub-arrays, each composed of five microphones. All microphones in a specific sub-array are characterized by the fact that they are co-planar. Each sub-array was devoted to the localization of the portion of reflectors lying on its plane. By intersecting line-reflectors estimated from multiple sub-arrays, the proposed methodology was shown to infer the actual lying plane of each reflector. Second, the approach in Section 4.2.3, that was proposed in [116], was extended to 3-D geometries. Additionally, the analytical prediction of the impact of errors on measurements on the reflector localization error was extended to the 3-D case [119].

Finally, in Chapter 6 a technique for inferring the geometry of a room under temperature variations was presented [171, 172]. Since existing methods rely on either synchronized measurements or often inaccurate estimates of the speed of sound, the propagation speed was estimated before the dominant reflectors were localized. In this way the room geometry could be reconstructed even when there are fluctuations in the ambient temperature. Experimental results in a real conference room, that was subject to strong variations in temperature with position, demonstrated improvements in accuracy by blindly inferring the actual speed of sound first. As a next step, possible extensions of the geometric inference framework were considered. It was shown that additional acoustic characteristics, such as the reverberation time or energy decay curves, could be inferred by taking into account the knowledge on the position of the reflectors in the acoustic environment.

7.1 Discussion

This thesis has demonstrated that the geometry of an acoustic enclosure, along with other related acoustic parameters, can be estimated using array processing techniques using minimum prior information. Rigorous theoretical developments along with real-world experiments demonstrated the high degree of accuracy achieved using the proposed approach.

In this work we considered first the solution of the wave equation, using the ray method, with low-frequency sound diffraction effects not taken into consideration. Under the assumption of specular reflections, the acoustic reflectors exhibit a mirror-like behaviour, motivating the derivation of a geometric constraint for localization. Using only three microphones in a 2-D plane (or correspondingly four microphones in 3-D) at arbitrary (yet known) positions inside a room, the location of acoustic reflectors was accurately estimated. Furthermore, no restrictions were placed on the input stimulus which could be a known signal (frequency sweep) or unknown source signal (hand clap or finger snap). This lead to the determination of the precise shape of a room with minimum prior assumptions. The repercussions of this work can potentially yield a paradigm change in architectural acoustics, echo cancellation and dereverberation, 3D sound reproduction and immersive reality applications.

Bibliography

[1] E. A. P. Habets and P. A. Naylor, "An online quasi-newton algorithm for blind SIMO identification," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Tx, Mar. 2010, pp. 2662–2665.

[2] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[3] F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Speech Audio Process.*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.

[4] JASA, "Physics and Astronomy Classification Scheme (PACS), section 43. Acoustics." [Online]. Available: http://asa.aip.org/jasa.html

[5] R. B. Lindsay, "The story of acoustics," *Journal Acoust. Soc. of America*, vol. 39, no. 4, pp. 629–644, 1966.

[6] E. Lippman, *Musical thought in ancient Greece*. Da Capo Press, 1975.

[7] H. Kuttruff, *Room Acoustics*, 4th ed. Taylor & Frances, 2000.

[8] B. Münxelhaus, *Pythagoras musicus: Zur Rezeption der pythagoreischen Musiktheorie als quadrivialer Wissenschaft im lateinischen Mittelalter*, ser. Orpheus-Schriftenreihe zu Grundfragen der Musik. Verlag für Systematische Musikwissenschaft, 1976.

[9] Galileo Galilei, H. Crew, and A. de Salvio, *Dialogues Concerning Two New Sciences*. Macmillan Company, 1914.

[10] L. M. A. Lenihan, "Mersenne and Gassendi. An Early Chapter in the History of Sound," *Acustica*, vol. 2, pp. 96–99, 1951.

[11] R. Gunther and R. Hooke, *The life and work of Robert Hooke, pt. 1-2*, ser. Early Science in Oxford. Dawsons of Pall Mall, 1930.

[12] J. Sauveur, *Principes d'acoustique et de musique: ou, Système général des intervalles des sons*. Minkoff Reprint, 1701.

[13] B. Taylor, "De Motu nervi tensi," *Phil. Trans. Roy. Soc.*, vol. 28, pp. 11–21, 1713.

[14] J. B. J. Fourier, *La Théorie analytique de la chaleur*. Paris: F. Didot, 1822.

[15] J. W. S. Rayleigh, *The theory of sound*. Macmillan and co., 1894, vol. 1.

[16] ——, *The theory of sound*. Macmillan and co., 1896, vol. 2.

[17] C. Truesdell and L. Euler, *The Rational Mechanics: Of Flexible Or Elastic Bodies*, ser. Leonhardi Euleri Opera Omnia. Orell Füssli, 1960.

[18] E. F. F. Chladni, *Entdeckungen über die Theorie des Klanges*. Leipzig: Weidmann, Erben & Reich, 1787.

[19] G. R. Kirchhoff, "Über das Gleichgewicht und die Bewegung einer elastischen Scheibe," *Journal für die reine und angewandte Mathematik*, vol. 40, pp. 51–88, 1850.

[20] F. V. Hunt, *Electroacoustics*. John Wiley & Sons Inc., 1954.

[21] W. Derham, "Experimenta et observationes de soni motu aliisque ad id attinentibus," *Phil. Trans. Roy. Soc.*, vol. 26, pp. 1–35, 1708.

[22] I. Newton, *Philosophiae naturalis principia mathematica*. J. Societatis Regiae ac Typis J. Streater, 1687.

[23] S. D. Poisson, "Sur le Mouvement des fluides élastiques dans les tuyaux cylindriques, et sur la théorie des instruments à vent," *Mém. Acad. Roy. Sci. Inst.*, vol. 2, p. 305, 1819.

[24] G. Green, "On the motion of waves in a variable canal of small depth and width," *Trans. Cambridge Phil. Soc.*, vol. 6, pp. 457–462, 1838.

[25] G. F. B. Riemann, "Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite," *Göttingen Abh.*, vol. 8, p. 43, 1860.

[26] S. Earnshaw, "On the Mathematical Theory of Sound," *Phil. Trans. Roy. Soc.*, vol. 150, p. 133, 1858.

[27] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*. New York: Interscience Publishers, Inc., 1948.

[28] S. E. Olive, "The Preservation of Timbre: Microphones, Loudspeakers, Sound Sources and Acoustical Spaces," in *Proc. 8th AES International Conference*, May 1990.

[29] R. Howe and M. Hawksford, "A methodological critique of local room equalization techniques," in *Digital Audio Signal Processing, IEE Colloquium on*, May 1991, pp. 6/1–6/4.

[30] W. C. Sabine, *Collected papers on acoustics*. Cambridge, MA: Harvard University Press, 1922.

[31] W. S. Franklin, "Derivation of equation of decaying sound in a room and definition of open window equivalent of absorbing power," *Phys. Rev. (Series I)*, vol. 16, pp. 372–374, Jun. 1903.

[32] C. F. Eyring, "Reverberation time in "dead" rooms," *Journal Acoust. Soc. of America*, vol. 1, no. 2A, pp. 168–168, 1930.

[33] J. R. Wright, "An exact model of acoustic radiation in enclosed spaces," *Journal Audio Eng. Soc.*, vol. 43, no. 10, pp. 813–820, Oct. 1995.

[34] B. M. Gibbs and D. K. Jones, "A simple image method for calculating the distribution of sound pressure levels within an enclosure," *Acustica*, vol. 26, pp. 24–32, 1972.

[35] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal Acoust. Soc. of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[36] J. Borish, "Extension of the image model to arbitray polyhedra," *Journal Audio Eng. Soc.*, vol. 75, pp. 1827–1836, 1984.

[37] U. R. Kristiansen, A. Krokstad, and T. Follestad, "Extending the image method to higher-order reflections," *Applied Acoustics*, vol. 38, pp. 195–206, 1993.

[38] L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*. New York: Wiley, 1962.

[39] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, Nov. 1968.

[40] E. De Geest and R. Garcea, "Simulation of room transmission functions using a triangular beam tracing computer model," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995, pp. 253–256.

[41] E. Deines, M. Betram, J. Mohring, J. Jegorovs, F. Michel, H. Hagen, and G. M. Nielson, "Comparative visualization for wave-based and geometric acoustics," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 1173–1180, Sep. 2006.

[42] G. B. Stan, J. J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, 2002.

[43] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *Journal Audio Eng. Soc.*, vol. 31, pp. 478–489, 1983.

[44] W. Chu, "Impulse-response and reverberation-decay measurements made by using a periodic pseudorandom sequence," *Applied Acoustics*, vol. 29, no. 3, pp. 193–205, 1990.

[45] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *Journal Acoust. Soc. of America*, vol. 97, pp. 1119–1123, 1995.

[46] E. Mommertz and S. Muller, "Measuring impulse responses with digitally pre-emphasized pseudorandom noise derived from maximum-length sequences," *Applied Acoustics*, vol. 44, no. 3, pp. 195–214, 1995.

[47] S. K. Olesen, J. Plogsties, P. Minnaar, F. Christensen, and H. Møller, "An improved MLS measurement system for acquiring room impulse responses," in *Proceedings of NORSIG 2000, IEEE Nordic Signal Processing Symposium*, 2000, pp. 117–120.

[48] R. Pompoli and N. Prodi, "Guidelines for acoustical measurements inside historical opera houses: Procedures and validation," *Journal of Sound and Vibration*, vol. 232, no. 1, pp. 281–301, 2000.

[49] K. B. Christensen, "The application of digital signal processing to large-scale simulation of room acoustics: Frequency response modeling and optimization software for a multi-channel dsp engine," *Journal Audio Eng. Soc.*, vol. 40, no. 4, pp. 260–276, Apr. 1992.

[50] K. Wyckaert, F. Augusztinovicz, and P. Sas, "Vibro-acoustical modal analysis: Reciprocity, model symmetry, and model validity," *Journal Acoust. Soc. of America*, vol. 100, no. 5, pp. 3172–3181, 1996.

[51] A. Kuntz and R. Rabenstein, "Room acoustic analysis: Complementing 2D measurements by 3D simulations," in *International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Mar. 2008, pp. 1492–1497.

[52] R. C. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *IEEE Signal Processing Society 12th DSP Workshop*, Jackson Lake, WY, Sep. 2006.

[53] ——, "Acoustical characterization of gunshots," in *IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, Apr. 2007, pp. 1–5.

[54] S. Uhlich and B. Yang, "Recursive estimation of room impulse responses with energy conservation constraints," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3648–3651.

[55] E. W. Taylor, "A preliminary study of the influence of rom mode structure on sound absorption," BBC, Tech. Rep. R&D Report 1983/4, 1983.

[56] R. Walker, "The design and application of modular, acoustic diffusing elements," BBC, Tech. Rep. R&D Report 1990/15, 1990.

[57] ——, "Optimum dimension ratios for studios, control rooms and listening rooms," BBC, Tech. Rep. R&D Report 1993/8, 1993.

[58] A. J. Berkhout, "A holographic approach to acoustic control," *Journal Audio Eng. Soc.*, vol. 36, no. 12, pp. 977–995, Dec. 1988.

[59] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[60] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 4, pp. 812–824, May 2008.

[61] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2009, pp. 313–316.

[62] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "A methodology for evaluating the accuracy of wave field rendering techniques," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 69–72.

[63] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, pp. 67–94, Jul. 1996.

[64] F. Talantzis, A. G. Constantinides, and L. C. Polymenakos, "Estimation of direction of arrival using information theory," *IEEE Signal Process. Lett.*, vol. 12, no. 8, pp. 561–564, Aug. 2005.

[65] W. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," *Journal Acoust. Soc. of America*, vol. 58, no. 1, pp. 201–207, Jul. 1975.

[66] B. Mungamuru and P. Aarabi, "Enhanced sound localization," in *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 3, Jun. 2004, pp. 1526–1540.

[67] M. W. Shan, T. and T. Kailath, "On spatial smoothing for direction-of-arrival estimation in coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 8, pp. 806–811, Aug. 1985.

[68] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[69] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[70] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, Piscataway, New Jersey, USA, Mar. 2005.

[71] ——, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Mar. 2005, pp. 89–92.

[72] K. W. Lo and C. K. Li, "An improved multiple target angle tracking algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 797–805, 1992.

[73] S. B. Park, C. S. Ryu, and K. K. Lee, "Multiple target angle tracking algorithm using predicted angles," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 643–648, 1994.

[74] C. R. Rao, C. R. Sastry, and B. Zhou, "Tracking the direction of arrival of multiple moving targets," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1133–1144, 1994.

[75] X. Gao, J. Filos, and W. Dai, "DOA Tracking via Simultaneous Angle-Source Update (SASU)," in *Sensor Signal Processing for Defence (SSPD 2012)*, Sep. 2012.

[76] ——, "DOA Tracking via Simultaneous Angle-Source Update (SASU)," *submitted to IEEE Trans. Wireless Commun.*, 2012.

[77] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 826–836, Nov. 2003.

[78] D. Simon, "From here to infinity," *Embedded Systems Programming*, pp. 2–9, Oct. 2001.

[79] Y. A. Huang, "Real-time acoustic source localization with passive microphone arrays," Ph.D. dissertation, Georgia Insitute of Technology, 2001.

[80] Y. A. Huang, J. Benesty, G. W. Elko, and R. M. Mersereati, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 943–956, Nov. 2001.

[81] B. T. Fang, "Simple solutions for hyperbolic and related fixes," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 5, pp. 748–753, Sep. 1990.

[82] R. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 8, no. 6, pp. 821–835, Nov. 1972.

[83] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[84] D. Hertz, "Time delay estimation by combining efficient algorithms and generalized cross-correlation methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 1, pp. 1–7, Feb. 1986.

[85] J. S. Abel and J. O. Smith, "Source range and depth estimation from multipath range difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1157–1165, Aug. 1989.

[86] B. Friedlander, "On the Cramer-Rao bound for time delay and doppler estimation," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 3, pp. 575–380, May 1984.

[87] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 1997.

[88] A. Cirillo, R. Parisi, and A. Uncini, "A new consistency measure for localization of sound sources in the presence of reverberation," in *International Conference on Digital Signal Processing*, Jul. 2009, pp. 1–5.

[89] A. Cirillo, R. Parisi, M. Scarpiniti, and A. Uncini, "Simplified optimal line selection for acoustic localization in the presence of reverberation," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1963–1967.

[90] J. O. Smith and J. S. Abel, "The spherical interpolation method for closed-form passive localization using range difference measurements," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, Dec. 1987, pp. 471–474.

[91] ——, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1661–1669, Dec. 1987.

[92] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 8, pp. 1223–1225, Aug. 1987.

[93] Y. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 2, no. 8, pp. 1905–1915, Aug. 1994.

[94] K. Cheung, H. So, W.-K. Ma, and Y. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1121–1130, Apr. 2004.

[95] B. Yang, "Different sensor placement strategies for tdoa based localization," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, Apr. 2007, pp. 1093–1096.

[96] Y. Lin and D. Lee, "Relevant deconvolution for acoustic source estimation," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, Mar. 2005, pp. 529–532.

[97] O. Lange and B. Yang, "Antenna geometry optimization for 2D direction-of-arrival estimation for radar imaging," in *Smart Antennas (WSA), 2011 International ITG Workshop on*, Feb. 2011, pp. 1–8.

[98] W. Pereira, D. Simpson, and J. Machado, "An estimator of focal position based on geometric acoustics," in *Ultrasonics Symposium, 1992. Proceedings.*, Oct. 1992, pp. 323–325.

[99] W. Tager and Y. Mahieux, "Reverberant sound field analysis using a microphone array," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Apr. 1997, pp. 383–386.

[100] R. Gang, G. Bocko, S. Roessner, J. Lundberg, D. Headlam, and M. Bocko, "Audio phase singularity detection for room acoustics parameter estimation," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2012, pp. 77–78.

[101] L. Del Pero, J. Guan, E. Brau, J. Schlecht, and K. Barnard, "Sampling bedrooms," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '11, Washington, DC, USA, 2011, pp. 2009–2016.

[102] A. O'Donovan and R. Duraiswami, "Microphone arrays as generalized cameras for integrated audio visual processing," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007.

[103] D. Aprea, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic reconstruction of the geometry of an environment through acquisitions of a controlled emission," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Glasgow, Scotland, Aug. 2009, pp. 710–714.

[104] D. Ba, F. Ribeiro, C. Zhang, and D. Florêncio, "L1 regularized room modeling with compact microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Tx, Mar. 2010, pp. 157–160.

[105] F. Ribeiro, D. Florêncio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 5, pp. 1449–1460, Jul. 2012.

[106] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 113–116.

[107] E. Mabande, H. Sun, K. Kowalczyk, and W. Kellermann, "On 2D localization of reflectors using robust beamforming techniques," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Prag, Czech Republic, May 2011, pp. 153–156.

[108] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Tx, Mar. 2010, pp. 153–156.

[109] S. Tervo and T. Tossavainen, "3D room geometry estimation from measured impulse responses," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[110] I. Dokmanic, Y. M. Lu, and M. Vetterli, "Can one hear the shape of a room: the 2–D polygonal case," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 321–324.

[111] I. Dokmanic, J. Ranieri, A. Chebira, and M. Vetterli, "Sensor networks for diffusion fields: Detection of sources in space and time," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Sep. 2011, pp. 1552–1558.

[112] I. Dokmanic and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 2617–2620.

[113] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Tx, Mar. 2010, pp. 2822–2825.

[114] J. Filos, E. A. P. Habets, and P. A. Naylor, "A two-step approach to blindly infer room geometries," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Sep. 2010.

[115] J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti, and P. A. Naylor, "Robust inference of room geometry from acoustic measurements using the Hough transform," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011, pp. 161–165.

[116] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor, and S. Tubaro, "Exact localization of acoustic reflectors from quadratic constraints," in *Proc. IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 17–20.

[117] E. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of planar acoustic reflectors through emission of controlled stimuli," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2011, pp. 156–160.

[118] J. Filos, A. Canclini, F. Antonacci, A. Sarti, and P. A. Naylor, "Localization of planar acoustic reflectors from the combination of linear estimates," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 1019–1023.

[119] A. Canclini, F. Antonacci, J. Filos, A. Sarti, and P. Naylor, "Exact localization of planar acoustic reflectors in three-dimensional geometries," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug., 2012, pp. 1–4.

[120] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.

[121] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[122] M. Thomas, N. Gaubitch, E. Habets, and P. Naylor, "An insight into common filtering in noisy SIMO blind system identification," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 521–524.

[123] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[124] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," in *Proc. IEEE*, vol. 86, no. 10, Oct. 1998, pp. 1951–1968.

[125] Y. Huang, J. Benesty, and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Processing*, vol. 6, no. 86, pp. 1278–1295, 2006.

[126] ——, "Adaptive blind multichannel identification," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2007, ch. 13, pp. 259–279, part B.

[127] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and newton algorithms for blind channel identification," *Signal Processing*, vol. 82, pp. 1127–1138, Aug. 2002.

[128] M. Haque and M. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 305–308, 2008.

[129] M. K. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, "Improving robustness of blind adaptive multichannel identification algorithms using constraints," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2005.

[130] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

[131] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661 – 672, Mar. 1996.

[132] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, oct. 2006.

[133] E. A. P. Habets. (2008, May) Room impulse response (RIR) generator. [Online]. Available: http://home.tiscali.nl/ehabets/rirgenerator.html

[134] B. D. Radlović, R. Williamson, and R. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 311–319, 2000.

[135] M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization - an overview," *Journal Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, Nov. 1993.

[136] A. Pietrzyk, "Computer modeling of the sound field in small rooms," in *Proc. AES Intl. Conf. on Audio, Acoustics & Small Spaces*, vol. 2, Copenhagen, Denmark, Oct. 1998, pp. 24–31.

[137] D. Botteldoore, "Finite-difference time–domain simulation of low-frequency room acoustic problems," *Journal Acoust. Soc. of America*, vol. 98, no. 6, pp. 3302–3308, 1995.

[138] A. Kulowski, "Error investigation for the ray tracing technique," *Applied Acoustics*, vol. 15, no. 4, pp. 263–274, 1982.

[139] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal Acoust. Soc. of America*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.

[140] J. Vanderkooy, "Aspects of MLS measuring systems," *Journal Audio Eng. Soc.*, vol. 42, pp. 219–231, 1994.

[141] C. Jung and R. Schramm, "Rectangle detection based on a windowed Hough transform," in *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*, Oct. 2004, pp. 113–120.

[142] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, pp. 11–15, January 1972.

[143] Q. Ji and R. M. Haralick, "Error propagation for the Hough transforms," *Pattern Recogn. Lett.*, vol. 22, no. 6-7, pp. 813–823, May 2001.

[144] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2001.

[145] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. Society for Industrial and Applied Mathematics (SIAM), 2005, p. 103.

[146] M. Loeve, "Probability theory," *Graduate Texts in Mathematics*, vol. 45, p. 46, 1963.

[147] N. Toronto, B. Morse, D. Ventura, and K. Seppi, "The Hough Transform's Implicit Bayesian Foundation," in *IEEE International Conference on Image Processing (ICIP)*, vol. 4, 2007, pp. 377–380.

[148] M. Brookes, P. A. Naylor, and J. Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. Speech Audio Process.*, vol. 14, 2006.

[149] D. M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," 1997, [Online] http://www.ee.imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

[150] G. Turin, "An introduction to matched filters," *IRE Transactions on Information Theory*, vol. 6, no. 3, pp. 311–329, Jun. 1960.

[151] L. R. Fincham, "Refinements in the impulse testing of loudspeakers," *Journal Audio Eng. Soc.*, vol. 33, no. 3, pp. 133–140, Mar. 1985.

[152] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770 – 1778, May 2008.

[153] R. Gilmore, *Catastrophe theory for scientists and engineers*. Dover Publications, Inc., 1993, reprint of the 1981 original.

[154] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[155] Y. Furukawa and Y. Shinagawa, "Accurate and robust line segment extraction by analyzing distribution around peaks in Hough spaces," *Comput. Vis. Image Underst.*, vol. 92, pp. 1–25, Oct. 2003.

[156] L. Rui and K. Ho, "Bias analysis of source localization using the maximum likelihood estimator," in *proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2012.

[157] B. Moran, S. D. Howard, and D. Cochran, "An information-geometric approach to sensor management," in *proc. of IEEE International Conference on Audio, Speeech and Language Processing, ICASSP*, 2012.

[158] S. Amari and H. Nagaoka, *Translations of mathematical monographs*. American Mathematical Society, 2000.

[159] A. Rosenfeld and I. Weiss, "A convex polygon is determined by its Hough transforms," *Pattern Recogn. Lett.*, vol. 16, no. 3, pp. 305–306, Mar. 1995.

[160] C.-H. Jeong and J.-G. Ih, "Reconstruction of sound source pressures in an enclosure using the phased beam tracing method," *Journal Acoust. Soc. of America*, vol. 126, no. 1, pp. 158–166, 2009.

[161] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, 2008, pp. 329–332.

[162] A. Sugiyama, A. Hirano, and K. Nakayama, "Acoustic echo cancellation for conference systems," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2004, pp. 17–20.

[163] T. Betlehem and T. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 117, pp. 2100–2111, 2005.

[164] F. Antonacci, D. Riva, G. Prandi, M. Tagliasacchi, A. Sarti, and S. Tubaro, "Efficient interferer cancellation based on geometrical information of the reverberant environment," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[165] P. Annibale and R. Rabenstein, "Accuracy of time-difference-of-arrival based source localization algorithms under temperature variations," in *Proc. of Int. Symposium on Communications, Control and Signal Processing, (ISCCSP), Limassol, Cyprus.* IEEE, 2010.

[166] Z. Dapeng, W. Baihai, X. Tibing, and L. Jianjun, "Study on relationship between sound wave and temperature of fault diagnose of liquid transmission system," in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA '09. International Conference on*, vol. 1, april 2009, pp. 728 –731.

[167] K. Huang, C. Huang, Y. Li, and M. Young, "A new method for temperature measurement in an infant incubator [using speed of sound]," in *[Engineering in Medicine and Biology, 1999. 21st Annual Conf. and the 1999 Annual Fall Meeting of the Biomedical Engineering Soc.] BMES/EMBS Conference, 1999. Proceedings of the First Joint*, vol. 2, oct 1999, p. 897 vol.2.

[168] G. Drysdale and R. Palmer, "Determining wind velocity and the speed of sound with redundant transponders for a spread spectrum acoustic ranging system," in *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, vol. 1, 2002, pp. 369 – 372 vol.1.

[169] P. Abbot, S. Celuzza, I. Dyer, B. Gomes, J. Fulford, and J. Lynch, "Effects of east china sea shallow-water environment on acoustic propagation," *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 2, pp. 192 – 211, april 2003.

[170] I. Jovanovic, L. Sbaiz, and M. Vetterli, "Acoustic tomography method for measuring temperature and wind velocity," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, may 2006.

[171] P. Annibale, J. Filos, P. Naylor, and R. Rabenstein, "Geometric inference of the room geometry under temperature variations," in *Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on*, Rome, Italy, May 2012, pp. 1–4.

[172] ——, "TDOA-based speed of sound estimation for air temperature and room geometry inference," *IEEE Trans. Speech Audio Process.*, vol. 21, no. 2, pp. 234–246, Feb. 2013.

[173] P. Annibale and R. Rabenstein, "Acoustic source localization and speed estimation based on Time-Differences-of-Arrival under temperature variations," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010.

[174] J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," *Journal Acoust. Soc. of America*, vol. 116, no. 2, pp. 958–969, Aug. 2004.

[175] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal Acoust. Soc. of America*, vol. 124, no. 1, pp. 269–277, 2008.

[176] B. Yang and M. Kreissig, "An introduction to consistent graphs and their signal processing applications," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2740–2743.

Declaration

I herewith declare that I have produced this manuscript without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This manuscript has not previously been presented in identical or similar form to any other examination board.

The thesis work was conducted from November 2008 to July 2012 under the supervision of Dr. Patrick A. Naylor at Imperial College London.

Jason Filos
London, U.K.