Politecnico di Milano
Dipartimento di Elettronica e Informazione
Dottorato di Ricerca in Ingegneria dell'Informazione

Elaborazione plenacustica nello spazio dei raggi: modellazione e analisi delle scene acustiche

Tesi di dottorato di:
Dejan Marković
ID: 754193

Relatore:
Prof. Augusto Sarti
Tutore:
Prof. Andrea Monti Guarnieri
Coordinatore del programma di dottorato:
Prof. Carlo Fiorini

XXV ciclo

Politecnico di Milano
Dipartimento di Elettronica e Informazione
Piazza Leonardo da Vinci 32 I 20133 — Milano

Politecnico di Milano
Dipartimento di Elettronica e Informazione
Dottorato di Ricerca in Ingegneria dell'Informazione

Plenacoustic processing in the ray space: applications to acoustic scene modeling and analysis

Doctoral Dissertation of:
Dejan Marković
ID: 754193

Advisor:
Prof. Augusto Sarti
Tutor:
Prof. Andrea Monti Guarnieri
Supervisor of the Doctoral Program:
Prof. Carlo Fiorini

XXV edition

Acknowledgments

I would like to thank my supervisor, Prof. Augusto Sarti, for giving me the opportunity, a lot of good ideas, support and encouragement during the last three years. I would also like to acknowledge the fine work of the other individuals who have contributed to this research and, first of all, Dr. Fabio Antonacci for his guidance, precious advices and patience. A very special thanks to my co-worker Antonio Canclini for help with challenges faced within the E.C. founded SCENIC project. I am also very grateful to Prof. Stefano Tubaro for his valuable advices.

I would also like to thank Prof. Walter Kellermann and his staff of the Friedrich-Alexander University in Erlangen for their kind hospitality and assistance during my stay and work at the LNT-LMS laboratory in Germany – it has been a pleasure to work with them. In addition, I would like to thank the reviewers of this thesis, Prof. Patrick Naylor (Imperial College London, UK) and Prof. Sergio Canazza (University of Padova, Italy), for their interest, very helpful suggestions and improvements.

I would like to thank all my wonderful friends and colleagues at ISPG/SMCLab: Paolo Bestagini, Massimiliano Zanoni, Daniele Valente, Andrea Galbiati, Simone Milani, Marco Marcon and all other past and present members of the lab. I greatly appreciated your company (and jokes) in this three years.

Finally, I would like to thank my family and, especially, my brother Bojan for patience and support during my studies.

Dejan

Sommario

Questa tesi affronta un insieme di problematiche attuali nell'ambito di elaborazione dei segnali audio tramite lo studio, rappresentazione, acquisizione/costruzione e uso delle *informazioni plenacustiche* che catturano la scena acustica "vista" da diversi puti dello spazio. In particolare, per la *modellazione della propagazione acustica*, l'informazione plenacustica prende la forma di *informazione di visibilità* che cattura le condizioni di visibilità di oggetti geometrici da diversi punti dello spazio. Questa informazione viene usata per una efficiente ed accurata simulazione della propagazione di onde sonore in ambienti complessi. Per quanto riguarda invece l'*analisi di scene acustiche*, l'informazione plenacoustica acquisita da una o più schiere di microfoni viene rappresentata in forma di un *immagine plenacustica* che cattura il campo sonoro proveniente da diverse direzioni sui diversi punti dello spazio. Questa immagine contiene un numero significativo di dettagli e informazioni sulla scena acustica. Seguendo le leggi dell'acustica geometrica, l'informazione plenacustica viene rappresentata in termini di raggi acustici in uno spazio parametrico indicato come lo *spazio dei raggi*. La parametrizzazione adottata per descrivere i raggi acustici permette un efficiente costruzione dell'informazione di visibilità ed una facile estrazione delle caratteristiche d'interesse dalle immagini plenacustiche. Regolarità e generalità della rappresentazione in termini di raggi acustici rende l'informazione plenacustica adatta ad una serie di possibili applicazioni.

Le applicazioni esaminate in ambito di questa tesi mostrano la validità dell'approccio proposto sia per quanto riguarda la modellazione della propagazione sonora, sia per quanto riguarda l'analisi di scene acustiche. In particolare, le applicazioni discusse seguono lo scenario applicativo in cui un sistema di riproduzione spaziale di campi acustici basato su una o più schiere di altoparlanti si trova ad operare

dentro un ambiente sconosciuto. Una sorgente sonora genera il campo acustico e le immagini plenacustiche catturate con una schiera di microfoni vengono esaminate per ricostruire la geometria dell'ambiente. Data l'informazione sulla geometria, i percorsi acustici che collegano la sorgente con la schiera vengono calcolati tramite una simulazione della propagazione di onde sonore. Il secondo algoritmo di analisi confronta i percorsi simulati con le acquisizioni e stima i coefficienti di riflessione di ogni singola parete riflettente presente nella scena acustica. Il modello geometrico del campo sonoro viene ottenuto sfruttando l'informazione sulla geometria e proprietà riflessive dell'ambiente ottenuti in precedenza. Infine, il sistema di riproduzione usa questo modello per generare il campo acustico desiderato compensando allo stesso tempo le riverberazioni indesiderate dell'ambiente che ospita il sistema.

Vengono analizzati i aspetti teorici, implementativi e le prestazioni statistiche dei algoritmi proposti. La validità del approccio proposto va, tuttavia, oltre le applicazioni discusse in questa tesi. Efficienza, regolarità e generalità della rappresentazione, insieme alla prospettiva di una sempre maggiore disponibilità di sistemi basati su dei microfoni/altoparlanti integrati a basso costo, rendono le idee sviluppate in questa tesi interessanti per una vasta gamma di possibili applicazioni tra cui caratterizzazione e separazione delle sorgenti, estrapolazione del campo, ecc.

Abstract

In this thesis we address a fairly broad range of current audio signal processing problems through the study, representation, acquisition/construction and use of the *plenacoustic data* that captures the acoustic scene as "seen" from different points in space. In particular, for the *modeling of acoustic propagation*, the plenacoustic data takes form of the *visibility information* that specifies the visibility of geometric objects from generic points in space. This information is used for an efficient and accurate simulation of acoustic propagation in complex environments. As far as the *analysis of acoustic scenes* is concerned, the plenacoustic data acquired by one or more microphone arrays is represented in form of *plenacoustic image* that captures the soundfield coming from a given direction at a given point in space. This image carries a great deal of information on the acoustic scene.

Following the laws of the geometrical acoustics, the plenacoustic data is represented in terms of acoustic rays in a space here referred to as the *ray space*. The adopted parameterization of the acoustic rays allows both an efficient construction of visibility information and an easy extraction of acoustic features from the acquired plenacoustic images. High regularity and generality make the ray space representation of plenacoustic data suitable for a variety of potential applications.

The applications examined in this dissertation show the validity of the proposed approach for purposes of both modeling of acoustic propagation and analysis of acoustic scenes. In particular, the examined applications follow a specific scenario in which an advanced spatial audio system aimed at reproducing the desired soundfield within a region of space is placed inside an unknown hosting environment. First, an acoustic source probes the environment and the plenacoustic images acquired by a microphone array are examined in order to infer the geometry of the environment.

The modeling engine computes the reflective paths between the source and the array. The modeled paths are then used by the second analysis algorithm that compares them with the acoustic measurements in order to estimate the reflection coefficients of all reflective surfaces in the environment. Given the information on geometry and reflection coefficients of the hosting environment, the modeling engine is used, once again, to model the sound propagation inside such environment. The rendering system uses this information to reproduce the desired soundfield by means of a loudspeaker array while compensating at the same time for the natural reverberation of the hosting environment.

Theoretical aspects, implementation issues and statistical performance of proposed algorithms are analysed. The validity of the proposed approach, however, is not limited just to the presented applications. Efficiency, regularity and generality of the representation, as well as the perspective of even further availability of inexpensive integrated microphone/loudspeaker arrays in near future, make the proposed tools attractive for a wider range of possible applications including source characterization and separation, wavefield extrapolation, etc.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Environment aware (space-time) audio processing is an emerging field of growing interest in the acoustic research community. An environment-aware system gathers information on the surrounding acoustic enclosure, and uses it to improve the performance of algorithms in specific applications, or to perform tasks that would not be possible otherwise. Under the hypotheses of validity of geometrical acoustics, this information gathering process turns out to be methodologically similar to that of image-based 3D scene reconstruction. We can therefore expect that further performance improvements in space-time audio processing systems would come from gathering information from a multitude of "acoustic viewpoints". This can be achieved through the cooperation between multiple spatially-distributed acquisition systems and/or by employing spatially extended microphone arrays.

In geometrical acoustics we approximate the soundfield with the superposition of acoustic beams ("bundles" of acoustic rays). This perspective needs be accommodated with a suitable mathematical representation. This is why, in this dissertation, we propose a novel parametric space (here referred to as "ray space") whose points represent acoustic rays, and discuss its impact on a variety of applications in space-time audio processing. In particular, we show how acoustic measurements are mapped onto the ray space and what consequences this brings. We show that this representation turns out to be very ef-

fective for collecting and organizing information on the acoustic scene (sources, reflectors, etc.), which simplifies the reconstruction process. In addition, the ray space turns out to be effective not just for analysis purposes. In fact, the same domain can be used for representing the visibility of geometric objects from generic points in space, which in turn can be exploited for efficiently modeling the acoustic propagation in complex environments, again in a fully geometric fashion.

In this first Chapter we introduce the motivations and the goals of this dissertation. This Chapter, in fact, represents a roadmap for the rest of the thesis and points out the connections, application scenarios and main ideas behind the developed algorithms.

1.1. Motivations

In modern multimedia applications, more and more frequently signal processing algorithms are required to work together towards an objective. Examples are the emerging audio signal processing techniques that exploit the knowledge of the world that surrounds them to boost their performance and pave the way for novel innovative applications. In the past few years a number of these space-time acoustic signal processing methodologies have appeared in literature [1]. We expect more and more of these algorithms to emerge in the near future as the knowledge of environment that hosts the acoustic system represents an important resource for efficiently addressing the challenging tasks of future multimedia applications. Environment awareness is a recurring theme of this work, which is addressed from two different points of view. In fact, environment awareness of acoustic systems means, in practice, two things:

- **Ability to observe the environment**, i.e. to become aware of the geometric (e.g. locations of acoustic sources and reflective surfaces) and radiometric (e.g. radiation patterns and reflection coefficients) properties of the environment that hosts the acoustic system and its own configuration within it;
- **Ability to adapt to the observed environment**, i.e. to use the information on the environment to extract valuable data that can be exploited to improve the performance of existing algorithms; to gather further insight on the impact of the system within the hosting environment; and to perform tasks that go beyond what is currently possible today (e.g. localization and tracking of sources that are beyond visibility).

These requirements introduce the two main topics addressed in this dissertation: *analysis of acoustic scenes* and *modeling of acoustic propagation*. While generally addressed as separate topics in the literature, in advanced applications effective solutions to such problems can be fruitfully employed in simultaneous, synergistic and cooperative fashion. Consider, for example, the following *application scenario*:

> The user purchases a new reproduction system based on a loudspeaker array for high-fidelity (home theater) sound reproduction. The system is installed in a room along with a number of other devices that incorporate a microphone array (e.g. TV, PC, game console, etc.). The reproduction system communicates with the other devices through a standardized interface; probes the scene with different acoustic stimuli; and collects the data captured by the sensing devices. The collected data is processed and the environment geometry is estimated. An integrated modeling engine uses the gathered geometric information to predict the reverberations coming from the different walls of the environment. The simulations of reflective paths are compared to the measurements in order to estimate the radiometric information of reflective surfaces. Finally, the geometric and radiometric information, along with the simulated soundfield, are used by the rendering engine to compensate and/or exploit effects of reverberations in order to generate the desired soundfield within the prescribed region of space. The result is the acoustic experience of a virtual environment that is not corrupted by the natural reverberation of the hosting environment.

In the near future we expect the above scenario and other similar ones to become possible in everyday environments, thanks to the introduction of low-cost, small-size sensors, emitters, signal handlers and processing platforms. In fact, the great challenge that we are called to face is to take space-time processing solutions outside of "controlled" laboratory environments. Simplicity, flexibility, and interactivity are becoming increasingly important in modern multimedia applications. This means that, in order to guarantee the desired level of interaction with the user, new systems must be designed to be easy to manage; autonomous; robust to noise, interference, estimation and modeling errors; and computationally efficient.

Traditional systems that employ a single microphone or small integrated arrays gather information about the environment as "seen" from just a single position in space that not necessary corresponds to the position (or area) of interest for a specific

application. This fact limits the reliability of the observation and, as a consequence, the robustness of such systems. In order to obtain as much information as they can about the acoustic scene and become truly environment aware, the acoustic systems should observe the environment from different positions in space, either combining the data from different, spatially distributed systems as suggested in the previous example, or employing a single spatially extended sensing system. However, in order to be useful, all the information gathered by the sensors (without losses) should be collected, organized and represented in the same space in which it can be displayed in such a way that all the acoustic objects of interest in the acoustic scene are easy to detect, recognize, discern and extract. All this needs to be implemented without giving up a certain flexibility and efficiency. The system should constantly monitor the environment, localize and track sources, rapidly update information and perform simulations of acoustic propagation that on their hand allow the system to become aware of the effects that its actions produce inside the given environment. These requirements are typical of interactive multimedia applications but tend to severely limit the range of possible solutions that can be adopted for this goal. Therefore, the success of future advanced audio signal processing systems will depend not only on the performance improvements of single algorithms but also on the integration between analysis and modeling algorithms and the degree of cooperation between different, spatially distributed systems.

1.2. Geometric representation of wavefields

At a first glance, the analysis of acoustic scenes and the modeling of acoustic propagation do not seem to share much in terms of scope and requirements. In real world scenarios, however, they often coexist within the same application and interact with each other. The representation that we need to adopt, must keep this need under consideration. We need a representation for the soundfield throughout space that is suitable for addressing environment-awareness issues and offers efficient tools for the development of multimedia applications. In other words, we need a wavefield model that is efficient, flexible, and treats geometric objects not simply as a boundary condition but more like functional elements. This is why we resort to geometrical acoustics [2], which (through the eikonal equation[1]) approximates the acoustic wavefield with acoustic rays whose direction is normal to the wave front.

[1] see Appendix A

In a homogeneous medium (e.g. in-air propagation), rays propagate as straight lines; originate from acoustic sources; are reflected by the reflective surfaces; and are collected by acoustic receivers.

Although not entirely physically accurate, the geometric (ray) representation of the soundfield is very accurate when the wavelength is much smaller than the size of geometric objects that interact with the wavefield, which is usually true when we consider reflections from walls of an environment. This is why geometric representations of wavefields are widely used in several areas. Typical examples are software applications for architectural acoustics. However, geometrical acoustics is also used for the analysis of acoustic scenes based on geometric constraints [3], [4], [5], [6], [7], in rendering applications [8], [9], [10], etc.

We begin with defining a suitable parameterization for the acoustic rays. The resulting parametric space of the acoustic rays is here referred to as *ray space*. The ray space representation of the acoustic wavefield, underlying both the developed modeling and the analysis frameworks, tries to link together and get the best of both worlds. As far as modeling is concerned, ray acoustics is widely used for geometric modeling [11], [12], [13], [14]. In this work, however, we go further by exploiting the effectiveness of the employed ray space parameterization to address performance issues of the modeling stage, in terms of both accuracy and computational burden. To this end we notice that during the propagation the acoustic rays bounce back and forth from geometric objects in the environment. As a consequence, getting an idea on how the different objects "see" each other can help predict how different rays will propagate in the environment. These *visibility* conditions are easily represented in the ray space. Therefore we address the modeling problem through the study, construction and use of the *visibility data structure* that efficiently encodes the visibility conditions between the geometric objects that compose the acoustic scenes.

The validity of the ray space representation of the acoustic wavefield goes well beyond the modelling purposes. The high regularity of the ray space representation of acoustic primitives makes it attractive for analysis purposes as well. Ray space represents all the acoustic primitives of the environment in terms of acoustic rays. A single sensing device measures the soundfield relative to just a portion of acoustic rays that propagate in the environment. Therefore, the different sensing systems (microphone arrays), placed in a different positions in the environment, capture different sets of acoustic rays. We could think to map these spatially distributed (*plenacoustic* [15]) measurements into the ray space, i.e. fill the corresponding re-

gions of the ray space with data obtained from the measurements, and than devise tools that exploit the regularity of the ray space to extract missing pieces of information. We call the ray space representation of the plenacoustic measurements the *plenacoustic image*. Then we formulate and address the analysis problems through the study, acquisition and use of the plenacoustic images that capture both geometric and radiometric properties of the acoustic scene.

Although used with different purposes, the visibility information and the plenacoustic image turn out to be closely related as they share the same domain. In fact, we will see that the visibility information used for the modeling of acoustic propagation in enclosures represents a special case of what we define as the ideal plenacoustic image. This link between the two tools, developed for modeling and analysis purposes respectively, is in accordance with the previously outlined application scenario. In the next paragraphs we present the main ideas behind the visibility approach to acoustic modeling and the plenacoustic representation of acoustic measurements; then we outline the main contributions of the dissertation and its structure.

1.3. Visibility information and its impact on modeling

Geometric methods for modeling of acoustic propagation are the most widespread in multimedia applications due to: their conceptual simplicity; good accuracy in predicting the early reverberations; and computational efficiency when compared to numerical methods [16, 17, 18] based on an approximate solution of the acoustic wave equation on a finite grid [19]. Geometric methods compute reflective paths as *rays* joining the *image sources* (wall-reflected versions of the acoustic source) and the receiver. This group of methods is represented mainly by: ray tracing; image source; beam tracing; and variations thereof. A short description of each method follows[2].

Ray tracing method [11, 20, 21, 22] casts a finite number of acoustic rays from the source; follows their propagation and interaction (reflections) with the environment; collects rays passing through/near the receiver.

Image source method [12, 23, 24] iteratively mirrors the acoustic source with respect to all reflectors that constitute the environment; a *visibility* test

[2] see Appendix B for more details

checks the feasibility of all image sources and the existence of acoustic paths joining them with the receiver.

Beam tracing method [25, 26, 27, 28, 13, 14, 29, 30, 31] groups rays in compact bundles (*beams*) originated from the same image source and falling to (illuminating) the same reflective surface; beams bounce off the reflective surfaces and split when the reflected bundle illuminates different reflectors; the *beam tree* data structure contains the branching relationship between acoustic beams; an efficient lookup procedure of the beam tree is used to find the paths joining the image sources and the receiver.

When compared to other geometric methods, the beam tracing method is the most general and the most efficient one in terms of accuracy and computational burden. In fact, the beam tracing technique avoids the spatial aliasing problem of the ray tracing method caused by the discrete sampling of rays (i.e. some important propagation paths could be missed by all sample rays). It also avoids the computationally-demanding visibility check of the image source method (i.e. all the image sources should be tested for *occlusion* by all the potential obstacles). In fact, the beam tracing procedure eliminates automatically, during the construction of the beam tree, the image sources that are never visible as the iterative mirroring of the source is done only with respect to reflectors that the beam encounters during the propagation. In order to better understand the generality of beam tracing with respect to ray tracing and image source, notice that both ray tracing and image source methods consist of a series of *visibility-along-a-line* checks (i.e. "what is first reflector that the ray meets?"; "does the ray reach the receiver?"). On the other hand the beam tracing method performs a number of *visibility-from-a-point* (image source) evaluations and encodes them in form of beams. As a consequence, the beam tree represents efficiently the visibility of the geometric space from the source position. While the construction of the beam tree is generally a computationally demanding task it does not depend on receiver position. Once the receiver position is specified the acoustic paths between image sources and the receiver (point-to-point visibility) represent a subset of the previously computed beam tree data structure and can thus be efficiently computed through a simple lookup procedure [14]. As we can expect, for static sources there is no need to re-compute the beam tree, therefore, unlike most other methods, the beam tracing is suitable for real-time operation when only the receiver is free to move. However, performance issues arise when the source is

moving as this requires the computation of a new beam tree. In complex environments there are usually many occlusions; occlusions cause beams to split and the beam tree to "branch out" numerous times. As a consequence, the computational effort required for the construction of the beam tree makes the beam tracing process suitable for real-time operation only for a limited set of simple geometries. On the other hand, the interactivity of acoustic systems means mainly the real-time operation of the modeling engine. At the same time, for the applications of interactive virtual acoustics (e.g. immersive gaming, walk-through architectural acoustics, etc.), the complexity of environments that we are required to deal with is increasing, moving further and further away from the classic "shoebox" type of room. In order to overcome these performance issues a number of optimization techniques were proposed in the literature. However, except for simple geometries, the computational effort of the beam tracing algorithm in general 3D environments remained unevenly divided between the computationally demanding construction of the beam tree and the efficient tracing of the acoustic paths.

One could wonder, therefore, if it is possible to generalize the procedure that makes the beam tracing more efficient that the ray tracing and the image source methods, or, more specifically:

> *Can we build a higher-order visibility data structure (visibility from a region) that does not depend on source position but only on environment geometry (can be evaluated in an offline phase) and makes it easy to compute the acoustic beams in a similar way the beam tree makes it easy to compute acoustics paths?*

The above question summarizes the idea behind the visibility-based beam tracing algorithm first presented in [32]. In [32] this goal was achieved for 2D environments and later extended to accommodate phenomena of diffraction and diffusion in [10]. However, as it will be clear later on, the generalization of this approach to 3D geometries was far from trivial, and it represents the first important contribution of this work.

Also notice that for the visibility-based beam tracing algorithm to work we need a suitable parametrization that allows:

- ▷ an efficient representation and computation of visibility information;
- ▷ an easy extraction of beams once the source position is specified.

To this end notice that the rays are used to specify the visibility of objects from different points in space, i.e.

Definition. *The object is visible from a given point in space if the ray departing from that point reaches the object without being occluded.*

Furthermore, the main primitive of all the geometric modeling techniques (including the beam tracing) is the acoustic ray and, thus, describing all the necessary information as a function of acoustic rays seems an obvious choice. Therefore, in order to compute the visibility information and trace acoustic beams as they propagate in the environment, the definition of the ray space is extremely useful.

1.4. Plenacoustic image – definition, acquisition and analysis

A number of acoustic measurements are performed in order to extract a great amount of information about the acoustic scene. For these measurements a number of sensors (microphones) and emitters (loudspeakers and other acoustic sources) are used. Acquired data is then processed by a variety of space-time algorithms in order to obtain the desired results. Intuitively, combining the measurements obtained with different systems in different positions in space can help improve the accuracy and robustness of the existing algorithms (e.g. choosing the best measurement for the given application or combining measurements in order to obtain more accurate information) and even perform task that could not be possible otherwise. In the near future, with the availability of inexpensive integrated microphone arrays, the collaboration between different, spatially distributed systems will become even more compelling and give rise to new interesting possibilities in acoustical signal processing. Sampling the soundfield from different positions in space, however, is just a first step towards these new opportunities. The acquired information on the viewed scene has to be collected and displayed in a space that shows a high degree of regularity, to be exploited by space-time signal processing algorithms.

As we sample the soundfield in different positions in space, the acquired samples trace the shape of a spatio-temporal function. If we can understand the properties of this function and its relation to the acoustic primitives that compose the acoustic scene we could fill the “gaps” and extract a great deal of information about the acoustic scene. The function that defines the sound pressure field in all positions in

space is known in literature as the *plenacoustic function* introduced in [15] as the acoustic counterpart of the plenoptic function [33]. Although the plenoptic function describes the optical wave field intensity as a function of position and direction (plus time and wavelength), in [15] the plenacoustic function was defined as the instantaneous acoustic pressure at given location without the directional information as the longer wavelengths of acoustic waves make it difficult to measure. However, as observed also in [15], this non-directional function can be turned into a directional one using phase information. We think of the plenacoustic function as a function of position and direction:

Definition. *The plenacoustic function specifies the (complex) amplitude of the acoustic wavefield impinging at a given position from a given direction.*

As a consequence, we represent the geometric objects of interest and the spatially distributed (plenacosutic) measurements in the parameter space of oriented lines, i.e. the ray space.

In this work we are interested in developing a "plenacoustic camera", which is able to take "snapshots" of the plenacoustic function and represent it in the ray space where its regularity can be exploited through signal processing. Unlike plenoptic (lightfield) cameras, in the acoustic domain we cannot rely on devices for capturing rays coming from a given direction. We will therefore use space-time processing instead, and cope with the resulting degradation with respect to the ideal plenacoustic function. In particular, to map the measurements into the ray space we will use the acoustic signals acquired with a number of small, compact microphone arrays. Using beamforming techniques, the arrays estimate the distribution of acoustic power through the computation of a number of pseudospectra (the output power of the beamformer for each look direction) from positions that are uniformly distributed over a line segment, which plays the role of "window of observation" of the acoustic scene. These pseudospectra are then collectively remapped onto the ray space. The ray space representation of (plen)acoustic measurements is similar to images in optics, where each pixel represents a ray passing through the pixel point and the optical center. However, in our representation the acoustic rays are not bound to pass through a single acoustic center and, therefore, we will call "plenacoustic image" the result of this mapping. The gray levels of this plenacoustic image are the values of the pseudospectra corresponding to the rays that cross the window of observation.

As the microphone arrays sample the window of observation they capture different portions of the ray space corresponding to different "views" of the acoustic scene. The high regularity of the ray space can then be used to extract missing pieces of information. We observe that the problems of environment inference, source characterization and separation, wavefield extrapolation, etc., could be formulated and addressed through the study of the plenacoustic images.

1.5. Plenacoustic data

At this point it is interesting to point out the relation between the plenacoustic image and the visibility information introduced before. The ray space acts as a basis around which we construct both the analysis and modeling frameworks. Therefore, the representation of the geometric primitives involved in the analysis and modeling processes is equivalent. The visibility information, constructed theoretically from the model of the environment, can be seen as a special case of ideal (theoretical) plenacoustic image in which the radiometric information has been ignored. In particular, the visibility information does not carry the information on the amplitude of acoustic rays, it just specifies the presence or absence of rays that encode the visibility of geometric objects. The ideal plenacoustic image, on the other hand, associate to each acoustic ray an amplitude in accordance to the radiometric information of acoustic primitives involved in the propagation process (e.g. radiation pattern, reflection coefficient). The real (acquired) pleanacoustic image is a distorted version of the ideal plenacoustic image due to non-ideal nature of the acquisition process and equipment. We will see that, as in the optical images, this non-ideality is given by sampling and "blurring". From the practical point of view the visibility information is a plenacoustic image with just two possible values: visible and not visible; or equivalently, the plenacoustic image is a "grayscale" version of the visibility information.

1.6. Main contributions

In this work we examine theoretical aspects, implementation issues and statistical performance of the algorithms for the construction, acquisition and use of the plenacoustic data intended here as both the visibility information and plenacoustic image. As the first simplified case of the plenacoustic image, the importance and

the general validity of the visibility data is here proven by:

1. reformulating the 2D visibility-based beam tracing algorithm in an oriented projective space;

2. generalizing the approach to general 3D environments.

The new parametrization of the 2D algorithm paves the way to its extension to general 3D geometries. In fact, the 3D algorithm requires the use of a particular type of coordinates defined in an oriented projective space [34]. Furthermore, the homogeneous parametrization of the 2D ray space makes it compatible with the current environment awareness methods such as [35] and [5]. The new parametrization brings also a twofold advantage: no normalization and the ability to distinguish rays with same direction but opposite orientation allow the representation of the visibility information to be reflector-independent; reflections become linear projective transformations (homographies). Generalization of the ray space and beam tracing to 3D represents an important contribution and, although defined mainly for 2D geometries, the major part of application algorithms presented in this work can be generalized for 3D environments using the new 3D ray space parametrization.

The use of the environment geometry for the construction of the visibility data structure can be seen as a model-driven approach to the use of the plenacoustic images (Figure 1.1 (a)). In particular, the applications we examine in this work are:

1. modeling of acoustic propagation in enclosures (beam tracing);

2. the use of the modeling engine to predict soundfields of both virtual environment, whose acoustic impression has to be reproduced, and real environment that host the rendering system, allowing the compensation of undesired reflections and the constructive use of reflective boundaries present in the real environment.

The second step is to reverse the paradigm and use the same domain for analysis purposes. Turning the problem around and defining the plenacoustic image as a powerful analysis instrument can be seen as a data-driven approach (Figure 1.1 (b)). Although the real (measured) pleanacoustic image represents a sampled and distorted version of the ideal (theoretical) plenacoustic image, it can be used for extracting a significant amount of information on the acoustic scene, as evidenced by a number of performed simulations and real world experiments. In particular, in this work we focus on:

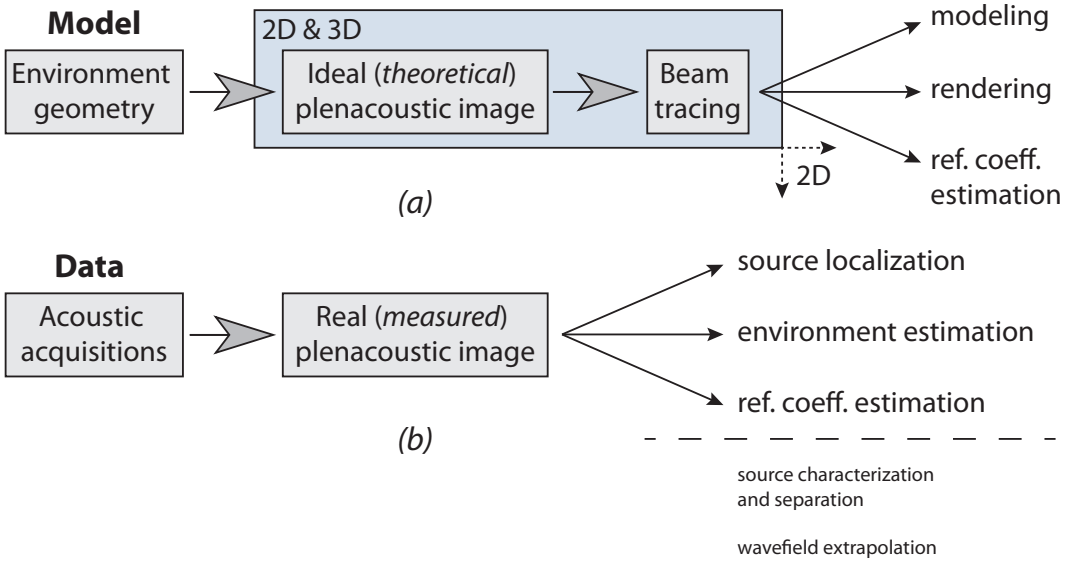


Figure 1.1.: (a) Overview of the model driven approach and its applications; (b) Overview of the data driven approach and its applications.

1. localization of acoustic sources;

2. environment geometry inference.

It will be shown that the proposed parametrization maps the data acquired with multiple observations in a space that is effective at collecting and displaying information on the acoustic scene as "seen" from different positions in space. Contributions of different objects can be fairly easily distinguished and extracted. The generality of the representation is evidenced by the fact that both source localization and geometry inference are performed in the same domain. The algorithms are easily extended for cases with multiple sources and reflectors. Furthermore, unlike the most part of inference algorithms, the proposed approach allows the estimation not only of the surface on which the reflector lies [36], [37], [38], [39], [3], [40]), [4], [41], [5] as typically addressed in the literature, but also of its extension in the case the edges are acoustically visible by the microphone array.

We also consider the problem of acoustic reflection coefficient estimation. We pursue this goal through the use of both analysis and modeling tools. In particular, the in-situ estimation of the acoustic reflection coefficients of reflective surfaces is performed through the matching between the measured pseudospectra, acquired by a microphone array, and the simulated pseudospectra, obtained using the modeling

engine. Although the reflection coefficient estimation is performed using the data obtained from a single position in space, this "simple" acoustic image can be seen as a narrow version of the more general plenacoustic image. Therefore, the reflection coefficient estimation algorithm developed in this thesis shows effectively that the data captured by the plenacoustic image contains not only geometric but also important radiometric information about the acoustic scene. Finally, although not part of this dissertation, we point out that the plenacoustic data can potentially be used also for source characterization and separation, and wavefield extrapolation. Currently the first works are carried out in this directions.

1.7. Structure of the thesis

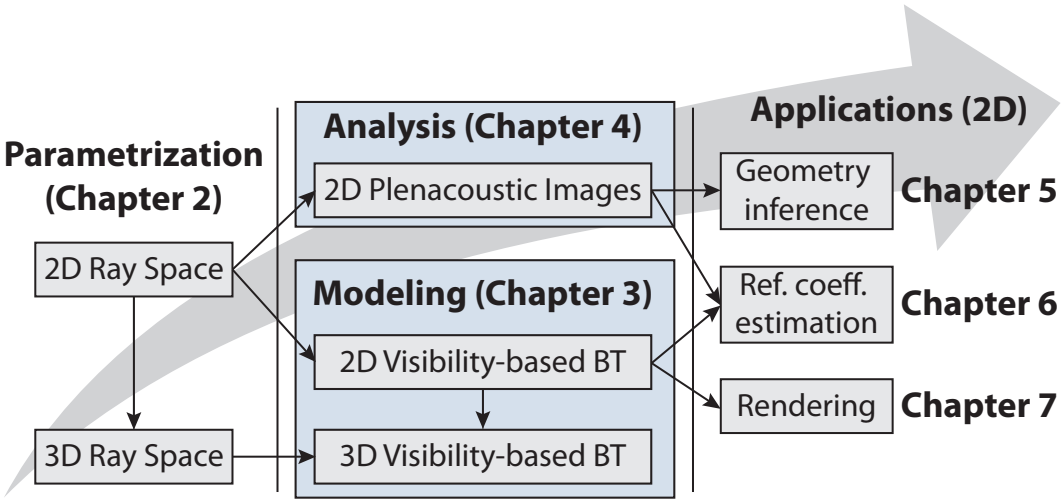


Figure 1.2.: Outline of the thesis.

An overview of the structure of this dissertation is given in Figure 1.2. This Figure is to be intended as a roadmap for the reader, showing connections and relations between the examined applications and developed tools. As an example the novel approach for the localization of reflecting boundaries and obstacles presented in Chapter 5 makes use of the concepts and tools presented in Chapters 2 and 4; theoretical aspects and issues encountered when implementing a fast acoustic propagation simulator are addressed in Chapters 2 and 3; etc. A brief description of each Chapter follows.

- **Chapter 2:** introduces both 2D and 3D parametrization of the ray space used throughout the rest of the work;
- **Chapter 3:** describes the developed 2D/3D visibility-based beam tracing algorithm and its performance;
- **Chapter 4:** examines the theoretical plenacoustic image and introduces its array-based approximation for the 2D configurations;
- **Chapter 5:** presents algorithms for localization of acoustic sources and reflective surfaces and analyzes their statistical performance (2D);
- **Chapter 6:** focuses on the estimation of reflective properties of walls (2D);
- **Chapter 7:** discusses the application of the modeling engine for rendering purposes (2D);
- **Chapter 8:** draws the conclusions and outlines possible future research directions.

Each Chapter starts with a short overview, motivations and a brief summary of the related works. The Chapters describe the representation (Chapter 2), construction (Chapter 3), acquisition (Chapter 4) and use (Chapters 5, 6, and 7) of the plenacoustic data. In particular, the applications addressed in Chapters 5, 6, and 7, represent the single components of the wider application scenario described at the beginning of this Chapter. All the Chapters are written to present and highlight the original contributions of the dissertation and contain only essential background information. For more details on the theoretical background the interested reader can refer to the following Appendixes:

- **Appendix A:** derives the wave equation and its hight-frequency approximation that stands behind all the geometric modeling techniques including the visibility-based beam tracing algorithm presented in the Chapter 3;
- **Appendix B:** reviews the most common geometric techniques for the modeling of acoustic propagation in enclosures and highlights their pros and cons;
- **Appendix C:** briefly describes the beamforming techniques used in Chapters 5 and 6;

- **Appendix D:** describes a beamshaping technique that can be used in the rendering application presented in the Chapter 7.

Chapter 2

Parameterization of acoustic rays

In this Chapter we define a space of oriented lines. As a generic point in this space corresponds to a ray in the geometric space it is here referred as the ray space. In particular, for the parametrization of the acoustic rays we resort to:

- **2D: homogeneous coefficients of lines;**
- **3D: Plucker coordinates of lines.**

Then we show how the geometric objects (more precisely their acoustical counterparts defined in terms of acoustic rays) "look like" in the newly defined spaces. In the next Chapters the ray space is used for the visibility evaluation, tracing of acoustic beams and mapping of the acoustic measurements.

2.1. Motivations

There are three main reasons for which we decided to represent the geometric primitives and the acoustic measurements in the parametric space of acoustic rays:

- ▷ geometric modeling techniques approximate wavefields with rays;
- ▷ rays are used for visibility computations;
- ▷ acoustic rays constitute the domain of the plenacosutic function.

2. Parameterization of acoustic rays

The eikonal equation[1] is a high-frequency approximation that substitutes the acoustic wavefield with acoustic rays. Therefore, the acoustic rays are used by the geometric modeling methods for the modeling of acoustic propagation. These rays interact with other geometric primitives during the propagation in an enclosure: they are originated from an acoustic source (e.g. a loudspeaker), they encounter obstacles (e.g. walls), reflect, and finally they are picked up by an acoustic receiver (e.g. a microphone).

Moreover, the rays are used to specify the visibility of objects from different points in space. In fact, two objects are visible if the line segment that joins them is not occluded by any other object in the environment, i.e. if the ray departing from the first object does not intersect any obstacle during the propagation before reaching the second object. In [32] a ray parametrization is used for the purpose of defining visibility diagrams that combine into a data structure that can be iteratively looked up for readily tracing beams of acoustic rays in enclosures. Visibility diagrams describe which reflectors are hit by each one of the rays that bounce off a prescribed (reference) reflector. As a reflected ray can only originate from an image source behind the reflector and pass through it, the visibility function of that reflector describes which reflectors can be seen when looking through the "window of observation" represented by the reflector.

The space of oriented lines (rays) represents the domain of the plenacoustic function as well. In fact, quite symmetrically to its optical counterpart, the plenacoustic function can be thought of as a parameterization of the sound field, which is a function that describes the sound pressure in every direction through every point in space. This means that, in the case of a 2D geometric domain, it can be written as a function $\mathcal{F}(\theta, x, y, t, f)$ of direction θ; position (x, y); time t; and frequency f. In this work the dependency on time and frequency will be kept hidden in order to simplify the notation. Furthermore, we can safely assume that the acoustic radiance remains constant from point to point along a ray, which reduces the degrees of freedom of this function to two. This is, of course, true in the absence of propagation losses due to absorption, etc. We recall that (in a homogeneous medium) an acoustic ray is an oriented line that is inherently perpendicular to the acoustic wavefront (i.e. collinear with the wave vector). A beam of acoustic rays originating from an acoustic source, therefore, identifies an infinite combination of infinitesimal planar wavefront contributions. In geometrical acoustics (just like in optical acoustics) we can rely

[1] see Appendix A

on the principle of conservation of radiance, which states that the acoustic radiance remains constant along the ray. This means that a suitable parameterization of the sound field is offered by the acoustic rays. This is a significant advantage, as ray-based representations have proven successful in various applications of soundfield analysis [3] and modeling [9], [32], [10].

Consequently, for *an efficient representation of plenacoustic measurements, an efficient encoding of visibility conditions, and an easy and elegant representation of interactions between geometric primitives*, a workable parametrization of acoustic rays is required.

2.2. Related work

In [32], for purposes of visibility computation and fast tracing of acoustic beams in 2D environments, a parametrization based on the location and the angular coefficient of the ray at the point of intersection with the reference reflector was used. Here, in order to define the 2D ray space, we use the projective paramaterization based on the parameters of the oriented lines that represent the acoustic rays. A similar parametrization was used for visibility computations in [42, 43]. The transition to an oriented projective space [34] allows us to simplify the representation of specular reflections and other transformations that are commonly encountered while dealing with the propagation modeling and geometry inference. In fact, we will see that no normalization is needed; the ability to distinguish rays with same direction but opposite orientations make the representation of the visibility information reflector-independent; and the reflections become linear projective transformations (homographies).

Furthermore, the new 2D parametrization paves the way for the extension to 3D geometries because the 3D algorithm requires the use of a particular type of coordinates defined in an oriented projective space. In fact, for the definition of the 3D ray space we resort to the Plucker coordinates of lines [44], [34], [45]. These coordinates have first been applied for visibility computations by Teller [46] to find lines passing through four other lines [47, 48], stab convex polygons [49], compute the antipenumbra of an light source [50], etc. In this work we use the Plucker coordinates to define the ray space representations of geometric primitives involved in the acoustic propagation and interaction with the environment: the acoustic rays; sources (e.g. loudspeakers); receivers (e.g. microphones); and reflectors (e.g. walls).

Although more complex than their 2D counterparts, these representations allow the same ideas of the 2D visibility-based beam tracing to be applied in the case of 3D geometries.

2.3. 2D Ray space

A ray can be seen as an oriented line in the geometric space. A line in 2D space is represented by a linear equation

$$l_1 x + l_2 y + l_3 = 0, \tag{2.1}$$

and can be parametrized with the line coefficients $[l_1, l_2, l_3]^T$. This parametrization defines a class of equivalence. In fact, the coordinates are homogeneous (scalable) as $\mathbf{l} = k[l_1, l_2, l_3]^T$, with $k \neq 0$, represent the same line. However, rays have a travel direction. In order to distinguish two rays lying on the same line (2.1) but with opposite orientations we limit the range of the scalar k to the positive or negative interval

$$\mathbf{l}_1 = k[l_1, l_2, l_3]^T, \ k > 0,$$

$$\mathbf{l}_2 = k[l_1, l_2, l_3]^T, \ k < 0.$$

Thus we define coordinates in an oriented projective space $\mathbb{P}^2$ [34]. As a generic point (l_1, l_2, l_3) corresponds to a ray in the geometric space, the Euclidean space $(\mathbb{R}^3)$ spanned by such homogeneous coordinates of lines is called the ray space (see Figure 2.1).

Just for clarity of visualization, from now on, we depict the ray space representations of geometric objects that constitute the acoustic scene in a reduced 2D ray space obtained by intersecting the ray space with a prescribed plane, as shown in Figure 2.1 (b). The parametrization in [32] corresponds to the reduced ray space obtained intersecting the ray space with a plane that excludes from the representation the rays parallel to the reflector whose visibility information is currently being evaluated, and does not discriminate the rays laying on the same lines but with opposite orientations. The representation in [32] requires the rotation and scaling of the whole geometric space with respect to each reflector in the environment. Instead, the new parametrization allows a reflector-independent representation. Despite the increase in the dimension of the ray space, the degrees-of-freedom, and therefore the

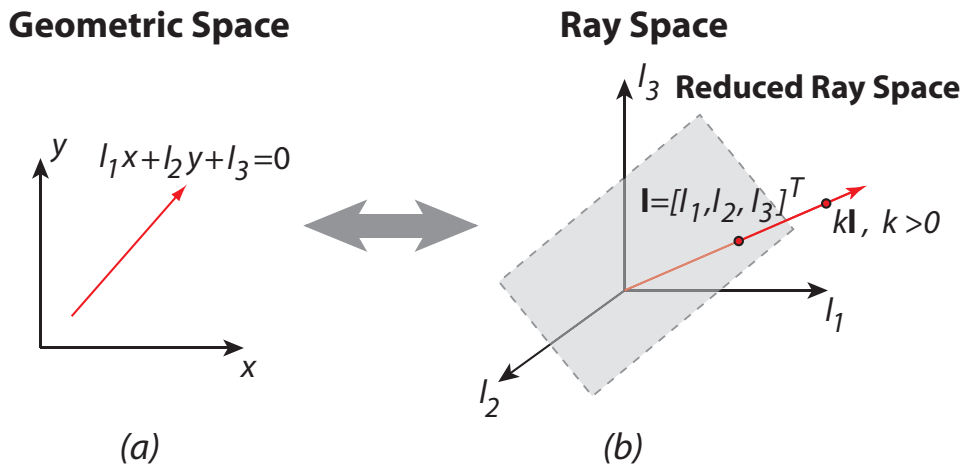


Figure 2.1.: (a) A ray in the geometric space; (b) In the ray space a ray corresponds to a projective point, here visualized as a line passing through the origin; the representation of the ray in the reduced ray space is obtained by slicing the ray space with the gray plane, yielding a point.

complexity of the problem, remains the same.

Geometric primitives – We are now interested in how the acoustic primitives involved in the propagation are represented in the ray space. The geometric primitives of interest are acoustic sources, receivers and reflectors. Sources and receivers are assumed to be point-like while reflectors are assumed to be planar. In order to represent the relations and interactions between the acoustic wavefield, represented by acoustic rays, and other geometric primitives, we have to map them into the ray space. We do this representing all the geometric primitives as sets of rays that intersect them.

2.3.1. Representation of a ray

As seen above we parametrize acoustic rays with (homogeneous) coefficients of the oriented lines they lie on. This parametrization defines the ray space. An acoustic ray in the geometric space corresponds to a half-line passing through the origin in the ray space, or a point in the reduced ray space, as shown in Figure 2.1.

Properties:

- Using the line equation (2.1), a ray $\mathbf{l}$ is passing through a point $\boldsymbol{x}_A = [x_A, y_A]^T$ with homogeneous coordinates $\mathbf{x}_A = k[x_A, y_A, 1]^T,\ k > 0$, if

$$\mathbf{x}_A^T \mathbf{l} = 0. \tag{2.2}$$

- Given two points, $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$, the two rays passing through them (going from $\boldsymbol{x}_A$ to $\boldsymbol{x}_B$ and from $\boldsymbol{x}_B$ to $\boldsymbol{x}_A$) are given by:

$$\begin{aligned} \mathbf{l}_{A\to B} &= \mathbf{x}_A \times \mathbf{x}_B, \\ \mathbf{l}_{B\to A} &= \mathbf{x}_B \times \mathbf{x}_A, \end{aligned} \tag{2.3}$$

  where $\times$ indicates the vector cross product. In fact, given the propriety of the cross product, $\mathbf{x}_A^T\mathbf{l}_{\mathbf{A}\to\mathbf{B}} = \mathbf{x}_B^T\mathbf{l}_{\mathbf{A}\to\mathbf{B}} = 0$ and $\mathbf{x}_A^T\mathbf{l}_{\mathbf{B}\to\mathbf{A}} = \mathbf{x}_B^T\mathbf{l}_{\mathbf{B}\to\mathbf{A}} = 0$.

2.3.2. Representation of sources and receivers

Acoustic sources and receivers can be seen as points in geometric space. A point is identified in the ray space by the set of all the rays that pass through it. Using the condition (2.2), the set of all rays passing through the point $\boldsymbol{x}_A$ with homogeneous coordinates $\mathbf{x}_A$ is

$$\overline{\mathbf{x}}_A = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} = 0\}. \tag{2.4}$$

Thus a point (source or receiver) in the geometric space corresponds to a plane passing through the origin in the ray space, or a line in the reduced ray space (see Figure 2.2 (a)).

Properties: Being a hyper-plane, $\overline{\mathbf{x}}_A$ divides the ray space into two half-spaces. This allows us to test the orientation of a ray with respect to the point as follows: all rays that have the point on their left or right, with respect to the travel direction, are given, respectively, by

$$\begin{aligned} \overline{\mathbf{x}}_A^+ &= \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} > 0\}, \\ \overline{\mathbf{x}}_A^- &= \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} < 0\}, \end{aligned} \tag{2.5}$$

as depicted in Figure 2.2 (b) and (c).

2.3.3. Representation of a reflector

In the geometric space the reflector $\boldsymbol{x}_A\boldsymbol{x}_B$ is a line segment completely defined by the two endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. In the ray space we represent it with the set of all rays that intersect this line segment. As each point of the line segment maps to a plane in the ray space (line in the reduced ray space), the reflector corresponds to the set of all the planes (lines) representing the infinite intermediate points found between

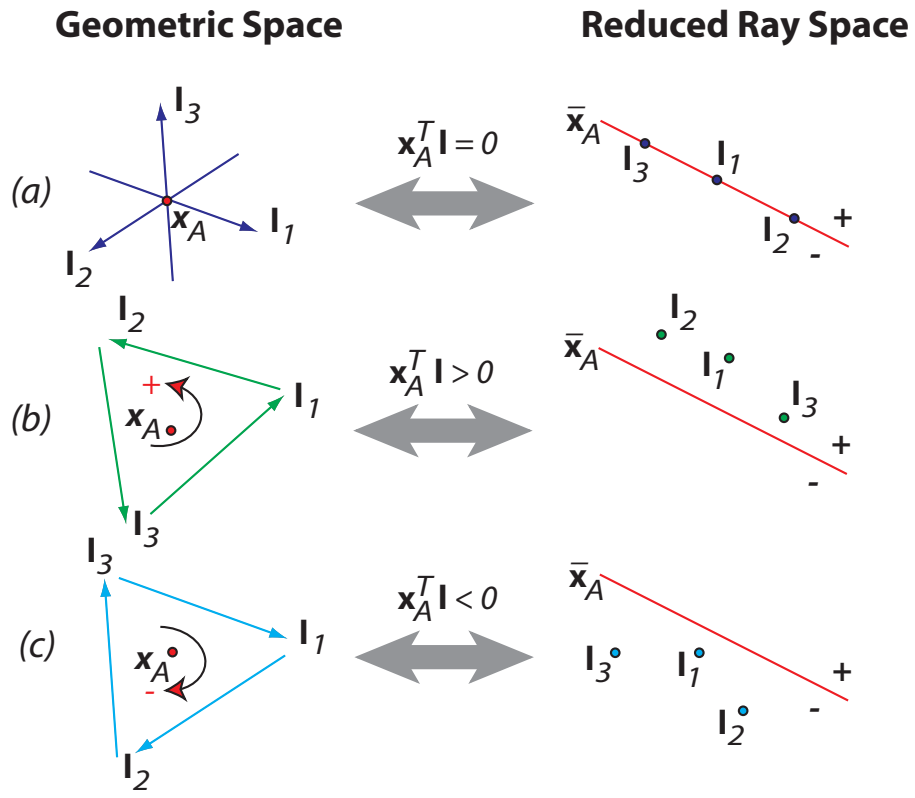


Figure 2.2.: The orientation of rays with respect to a point in the geometric space and their configuration in the ray space: (a) rays passing through the point $\boldsymbol{x}_A$ in the geometric space lie on the hyperplane $\overline{\mathbf{x}}_A$ in the ray space; (b) rays that travel counterclockwise with respect to the point $\boldsymbol{x}_A$ in the geometric space lie in the half-space $\overline{\mathbf{x}}_A^+$ in the ray space; (c) rays that travel clockwise with respect to the point $\boldsymbol{x}_A$ in the geometric space lie in the half-space $\overline{\mathbf{x}}_A^-$ in the ray space.

the two endpoints. Exploiting the orientation relations (2.5), in a similar way we defined two rays for the same line, we can distinguish two oriented reflectors $\overline{R}_{AB}^{(1)}$ and $\overline{R}_{AB}^{(2)}$, i.e. the two reflectors defined by the same line segment but characterized by different directions of incident rays

$$\begin{aligned}\overline{R}_{AB}^{(1)} &= \overline{\mathbf{x}}_A^- \cap \overline{\mathbf{x}}_B^+,\\ \overline{R}_{AB}^{(2)} &= \overline{\mathbf{x}}_A^+ \cap \overline{\mathbf{x}}_B^-,\end{aligned} \tag{2.6}$$

as illustrated in see Figure 2.3. The definition of an oriented reflector will turn useful during the evaluation of the visibility information. In fact, in accordance with the image source principle, when we evaluate the visibility of the environment from a mirrored image source, we do not consider the reflectors in the half-space where

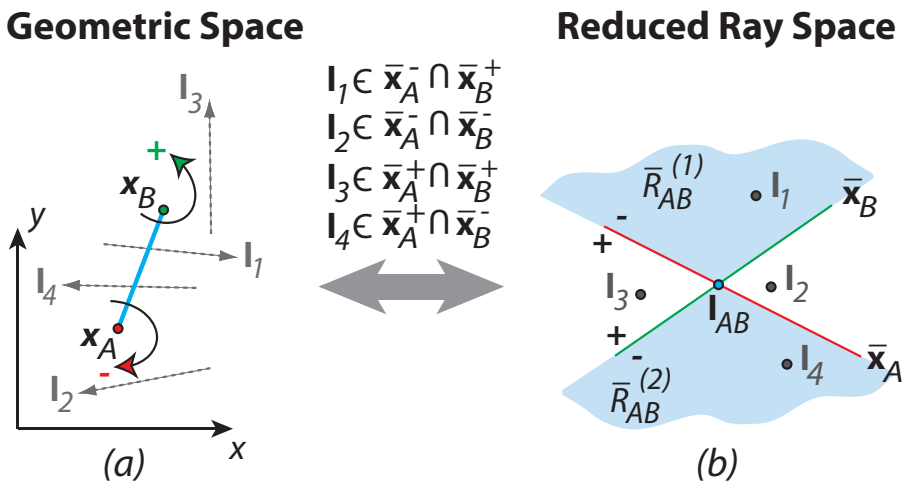


Figure 2.3.: (a) A segment in geometric space identifies four classes of rays: those that cross the segment in one direction ($\mathbf{l}_1$), those the cross it in the opposite direction ($\mathbf{l}_4$), those that do not cross the segment and are oriented in accordance to the first endpoint ($\mathbf{l}_2$) and those that do not cross the segment and are oriented in accordance to the second endpoint ($\mathbf{l}_3$); (b) The regions of the reduced ray space corresponding to such classes of rays are four quadrants given by the intersection of four half-spaces defined by the two endpoints; the quadrants originate from the point ($\mathbf{l}_{AB}$), which identifies the line that the segment lies on.

the mirrored source lies. Therefore, the use of an oriented reflector is necessary for correct evaluation of visibility conditions from such a mirrored image source.

Finally, the non oriented reflector $\overline{R}_{AB}$ can be expressed in a closed form as the union of two oriented reflectors that compose it, i.e.

$$\overline{R}_{AB} = \overline{R}_{AB}^{(1)} \cup \overline{R}_{AB}^{(2)}. \tag{2.7}$$

Therefore, the segment $\boldsymbol{x}_A\boldsymbol{x}_B$ is represented in the ray space as the set of rays $\overline{R}_{AB}$, i.e. the intermediate region between the hyperplanes $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$, representations of $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$, respectively.

2.3.4. Representation of a beam

Although not a real geometric primitive but rather a combination of different primitives, we present here the ray space representation of the acoustic beam. A beam is a compact bundle of rays that originate from an acoustic source and fall onto a surface (reflector). It is completely specified by an origin (source) and by the (connected) illuminated region of the surface. The surface can be completely illuminated by a beam (Figure 2.4 (a)) or partially, as it is occluded (Figure 2.4

(b)). In both cases it corresponds in the ray space to a part of a source hyperplane $\overline{\mathbf{x}}_S$ limited by two points as shown in Figure 2.4 (c).

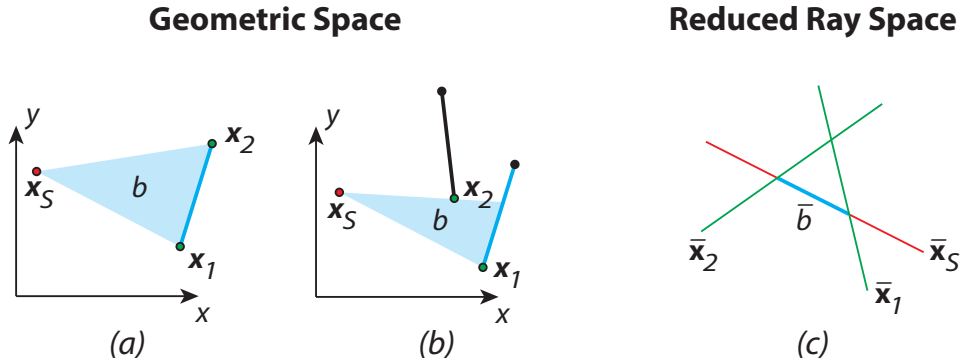


Figure 2.4.: (a)(b) Examples of a beam in the geometric space; (c) The beam in the ray space.

2.4. 3D Ray space

Extending the representation of geometric primitives from 2D to 3D space is not a trivial task. In 2D we used rays to represent all geometric primitives in the ray space. In 3D we would like to do the same. However, while the increase in dimension of the geometric space is 1, the degrees-of-freedom in the description of a line (ray) increase from 2 to 4. Being a flat subset with dimension 1, a line in a three-dimensional space is not a hyperplane, i.e. it does not separate the geometric space in two half-spaces as it does in 2D space (the 3D equivalent of a 2D line is a plane). As a consequence the representation of rays in 3D becomes more difficult. Furthermore, the visual events are complex in 3D: the discontinuity surfaces in visibility (e.g. boundaries between total, partial and no visibility) are not necessarily linear nor convex. Similar considerations are made in [51] and [52].

2.4.1. The ray space revisited

As in the 2D case we represent rays as oriented lines. The orientation of lines helps us define other geometric primitives involved in the acoustic propagation.

We can represent rays in 3D using the Plucker coordinates of lines [44], [34], [45]. Given two points with Cartesian coordinates $\boldsymbol{x}_A = [x_A, y_A, z_A]^T$ and $\boldsymbol{x}_B = [x_B, y_B, z_B]^T$, the Plucker coordinates of two oriented lines (rays) passing through

these points are

$$\mathbf{l}_1 = k \begin{bmatrix} \boldsymbol{x}_B - \boldsymbol{x}_A \\ \boldsymbol{x}_A \times \boldsymbol{x}_B \end{bmatrix}, \; k > 0,$$

$$\mathbf{l}_2 = k \begin{bmatrix} \boldsymbol{x}_A - \boldsymbol{x}_B \\ \boldsymbol{x}_B \times \boldsymbol{x}_A \end{bmatrix}, \; k > 0.$$

Thus the rays are defined in a five-dimensional oriented projective space $\mathbb{P}^5$. Furthermore, from the above definition a ray is given by

$$\mathbf{l} = \begin{bmatrix} \mathbf{l}^d \\ \mathbf{l}^m \end{bmatrix}, \; s.t. \; (\mathbf{l}^d)^T \mathbf{l}^m = 0, \tag{2.8}$$

where $\mathbf{l}^d$ is the vector given by the difference between point positions $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ (displacement or directional part) and $\mathbf{l}^m$ is the vector given by their vector product (moment or locational part). As a consequence not every point in $\mathbb{P}^5$ correspond to a ray in $\mathbb{R}^3$, only those points that satisfy the constraint $(\mathbf{l}^d)^T \mathbf{l}^m = 0$, i.e. only those points that lie on a four-dimensional ruled quadric surface known as the Plucker quadric

$$\mathbf{l}^T \mathbf{Q} \mathbf{l} = 0, \; \mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{I}_3 & \mathbf{0} \end{bmatrix}, \tag{2.9}$$

where $\mathbf{I}_3$ is 3×3 identity matrix.

The Plucker quadric embedded in $\mathbb{P}^5$ represents the ray space for rays in 3D geometry. Note, however, that it is not possible to graphically visualize the four-dimensional quadric in a five-dimensional projective space. Therefore, from now on, to give an idea of configurations in the ray space we make an illustration of two-dimensional ruled quadric in a three-dimensional Euclidean space as depicted in Figure 2.5.

2.4.2. Representation of a ray

As described above we parametrize rays with six coordinates as in (2.8). These are homogeneous coordinates of a point in $\mathbb{P}^5$ lying on the quadric surface (2.9) (see Figure 2.5).

The second interpretation of Plucker coordinates is also possible. We can represent

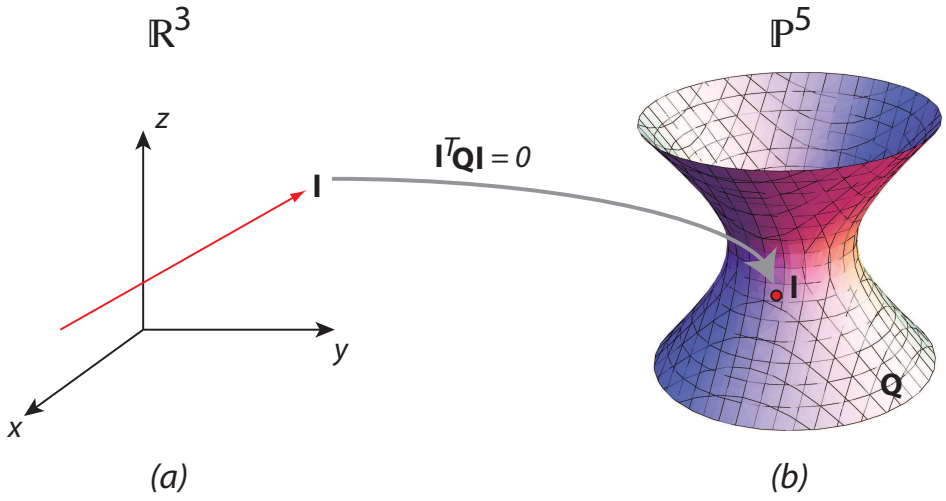


Figure 2.5.: (a) A ray in the geometric space; (b) A ray in the ray space.

a ray, permuting the order of coordinates, as

$$\widetilde{\mathbf{l}} = \begin{bmatrix} \mathbf{l}^m \\ \mathbf{l}^d \end{bmatrix}, \tag{2.10}$$

and interpret it as a hyperplane in $\mathbb{P}^5$. The use of these two representations, (2.8) and (2.10), allows us to represent the relative orientations of rays as shown in Figure 2.6. We can test the orientation of rays, represented in $\mathbb{P}^5$ using (2.8), with respect to a ray $\mathbf{l}_1$, represented in $\mathbb{P}^5$ using (2.10), by testing on which side of the hyperplane $\widetilde{\mathbf{l}}_1$ they are found. Rays that lie on $\widetilde{\mathbf{l}}_1$ in $\mathbb{P}^5$ intersect (or are parallel, i.e. intersect at infinity) the ray $\mathbf{l}_1$ in $\mathbb{R}^3$. With this consideration at hand it is interesting to observe that the Plucker condition (2.9) states simply that a ray intersects itself, i.e. $\mathbf{l}^T \mathbf{Q} \mathbf{l} = \widetilde{\mathbf{l}}^T \mathbf{l} = 0$.

2.4.3. Representation of a reflector

In a similar way we defined an oriented reflector in 2D with constraints on a correct orientation of rays with respect to reflector's endpoints, we define a reflector in 3D with constraints on a correct orientation of rays with respect to reflector's edge lines. A comparison between the representations of a reflector in 2D (depicted this time in the complete ray space) and a reflector in 3D (illustrated in a simplified 3D

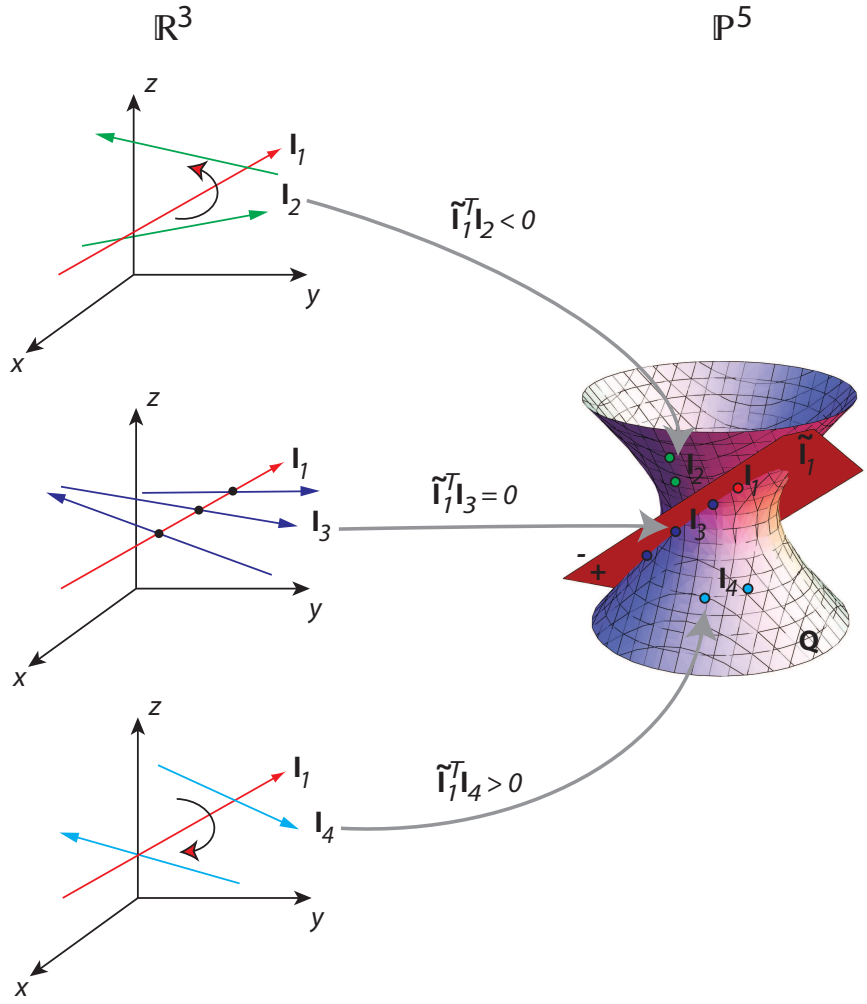


Figure 2.6.: Orientation of rays in the geometric space and their configuration in the ray space.

representation of the ray space) is shown in Figure 2.7. While the reflector in 2D is always defined by two endpoints, the reflector in 3D can be defined by an arbitrary number N of edges $\mathbf{e}_i$ (three for a reflector shown in Figure 2.7 (b)). Furthermore, in 3D one more constraint has to be added: the Plucker condition. As a consequence the reflector in 3D space is given in $\mathbb{P}^5$ by the intersection of N half-spaces defined by hyperplanes $\tilde{\mathbf{e}}_i$ (this ensures that rays that define the reflector have the right orientation with respect to edge rays $\mathbf{e}_i$, $i = 1, ...N$) with the Plucker quadric (this

ensures that points in $\mathbb{P}^5$ correspond to real rays in $\mathbb{R}^3$), i.e.

$$\overline{R} = \{\mathbf{l} \in \mathbb{P}^5 | \tilde{\mathbf{e}}_i^T \mathbf{l} \rhd_i 0, i = 1, ..., N, \mathbf{l}^T \mathbf{Q} \mathbf{l} = 0\}, \tag{2.11}$$

where "$\rhd_i$" can represent either ">" or "<" depending on the orientation constraint with respect to the edge $\mathbf{e}_i$.

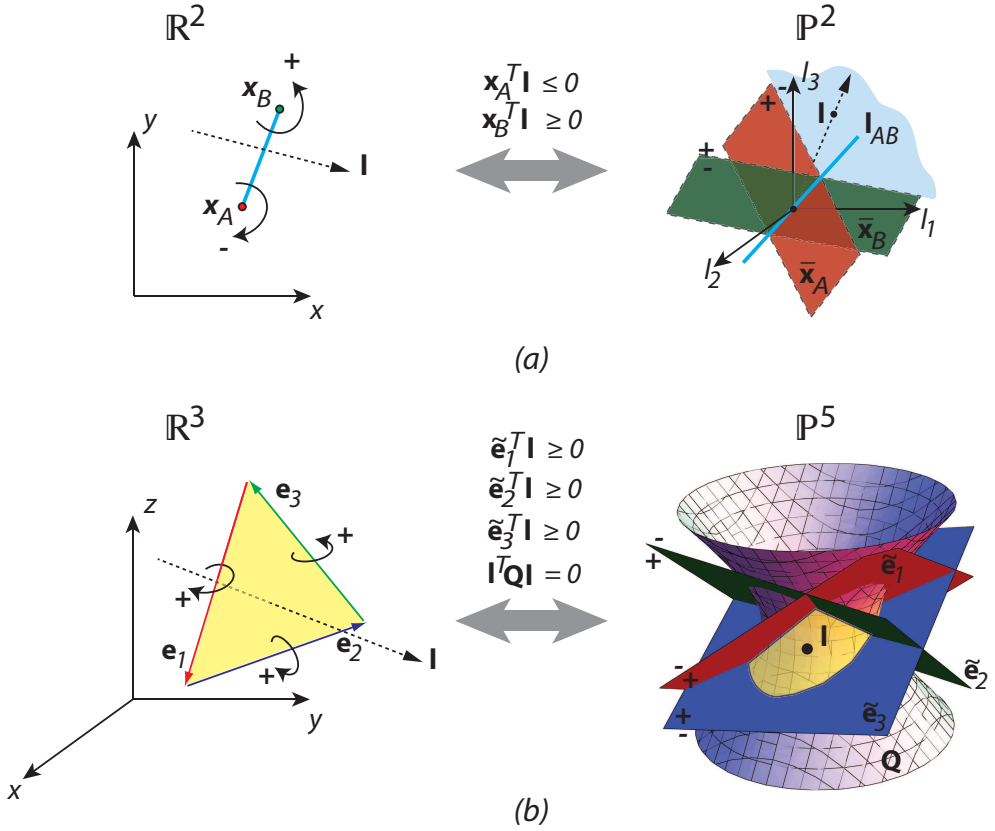


Figure 2.7.: (a) Reflector in 2D and its representation in the ray space; (b) Reflector in 3D and its representation in the ray space.

2.4.4. Representation of sources and receivers

Point-like sources and receivers are identified, once more, by the set of all rays that pass through them. However, given a point with Cartesian coordinates $\boldsymbol{x}_P = [x_P, y_P, z_P]^T$, it is not straightforward to constrain a generic line $\mathbf{l}$, represented using Plucker coordinates, to pas trough $\boldsymbol{x}_P$, as it was in 2D case. The point $\boldsymbol{x}_P$ and the

line $\mathbf{l}$ are defined in two different spaces, and therefore they cant "speak" to each other directly. However, we can constrain a ray $\mathbf{l}$ to pass through $\boldsymbol{x}_P$ by constraining it to intersect three non-coplanar lines, $\mathbf{l}_1$ $\mathbf{l}_2$, $\mathbf{l}_3$, all passing through the point $\boldsymbol{x}_P$, i.e.

$$\widetilde{\mathbf{l}}_1^T\mathbf{l} = 0, \quad \widetilde{\mathbf{l}}_2^T\mathbf{l} = 0, \quad \widetilde{\mathbf{l}}_3^T\mathbf{l} = 0.$$

As a consequence the ray $\mathbf{l}$ belongs to the null space of $\mathbf{L} = [\widetilde{\mathbf{l}}_1, \widetilde{\mathbf{l}}_2, \widetilde{\mathbf{l}}_3]^T$. Choosing three perpendicular lines, each parallel to one main axe (x, y and z)

$$\mathbf{l}_1 = \begin{bmatrix} \boldsymbol{x}_1 - \boldsymbol{x}_P \\ \boldsymbol{x}_P \times \boldsymbol{x}_1 \end{bmatrix} \quad \mathbf{l}_2 = \begin{bmatrix} \boldsymbol{x}_2 - \boldsymbol{x}_P \\ \boldsymbol{x}_P \times \boldsymbol{x}_2 \end{bmatrix} \quad \mathbf{l}_3 = \begin{bmatrix} \boldsymbol{x}_3 - \boldsymbol{x}_P \\ \boldsymbol{x}_P \times \boldsymbol{x}_3 \end{bmatrix},$$

where $\boldsymbol{x}_1 = [x_P + 1, y_P, z_P]^T$, $\boldsymbol{x}_2 = [x_P, y_P + 1, z_P]^T$, and $\boldsymbol{x}_3 = [x_P, y_P, z_P + 1]^T$, we have

$$\mathbf{L} = \begin{bmatrix} 0 & z_P & -y_P & 1 & 0 & 0 \\ -z_P & 0 & x_P & 0 & 1 & 0 \\ y_P & -x_P & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Supposing $x_P, y_P, z_P \neq 0$, a basis for the null space of $\mathbf{L}$ is given by

$$\mathbf{U}_P = \begin{bmatrix} 0 & 1/z_P & x_P/z_P \\ 1/x_P & y_P/(x_P z_P) & y_P/z_P \\ 0 & 0 & 1 \\ -z_P/x_P & -y_P/x_P & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{2.12}$$

Note that an equivalent representation can be found in case above conditions on x_P, y_P, z_P are not met. The ray $\mathbf{l}$ belongs to the vector space spanned by vectors $\mathbf{U}_P = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$, i.e. it can be written as a linear combination

$$\mathbf{l} = \mathbf{u}_1 a_1 + \mathbf{u}_2 a_2 + \mathbf{u}_3 a_3 = \mathbf{U}_P \mathbf{a}, \tag{2.13}$$

where $\mathbf{a} = [a_1, a_2, a_3]^T$. Notice that $\mathbf{U}_P(k\mathbf{a}) = k(\mathbf{U}_P\mathbf{a}) = k\mathbf{l}$, i.e. the coefficients $\mathbf{a}$ are three homogeneous coordinates that specify a point on the projective plane $\mathbb{P}^2$.

Furthermore, from (2.12) it is not difficult to verify that

$$\begin{aligned}\mathbf{u}_i^T\mathbf{Q}\mathbf{u}_j = 0,\ i,j = 1,2,3,\\ \Longrightarrow \mathbf{l}^T\mathbf{Q}\mathbf{l} = \textstyle\sum_{i=1}^{3}\sum_{j=1}^{3} a_i a_j \mathbf{u}_i^T\mathbf{Q}\mathbf{u}_j = 0,\end{aligned}$$

or, equivalently, each ray $\mathbf{l}$ given by (2.13) lies on the surface of the Plucker quadric (2.9).

Finally we can state that the point $\boldsymbol{x}_P$ (acoustic source or receiver) is given by the set of all rays that pass through it, i.e. by the plane (spanned by (2.12)) that lies completely on the surface of the (ruled) Plucker quadric embedded in $\mathbb{P}^5$, i.e.

$$\overline{\mathbf{x}}_P = \{\mathbf{l} \in \mathbb{P}^5 | \mathbf{l} = \mathbf{U}_P\mathbf{a}, \mathbf{a} \in \mathbb{P}^2\}. \tag{2.14}$$

2.4.5. Representation of a beam

We recall that a beam is a compact bundle of rays that originate from an acoustic source and fall onto a reflector. In the ray space it is represented by a portion of the source's projective plane (2.14) limited by a number of hyperplanes corresponding to edges $\mathbf{e}_i$ that limit the beam (reflector edges and/or edges that limit the visibility of the reflector). When compared to the beam in 2D the increase in dimensionality is just one despite the ray space is defined in $\mathbb{P}^5$ instead of $\mathbb{P}^2$ (recall that in 2D the beam was a projective line limited by two points).

For purpose of determining the edges that limit the beam we can reduce the complexity (dimensionality) of the task by projecting the hyperplanes $\widetilde{\mathbf{e}}_i$ from $\mathbb{P}^5$ onto the projective plane $\mathbb{P}^2$ defined by coefficients $\mathbf{a}$ that specify lines passing through the source (see (2.12) and (2.13)). The projection is performed as follows

$$\begin{aligned} g_P &: \mathbb{P}^5 \to \mathbb{P}^2 \\ \bar{\mathbf{e}}_i &= g_P(\widetilde{\mathbf{e}}_i) = \mathbf{U}_P^T\widetilde{\mathbf{e}}_i, \end{aligned} \tag{2.15}$$

where $\mathbf{U}_P$ is the matrix in equation (2.12). Using the coefficients $\mathbf{a}$ and projections $\bar{\mathbf{e}}_i$ we can still test the orientation of the ray $\mathbf{l}$ with respect to the edge $\mathbf{e}_i$, in fact $\widetilde{\mathbf{e}}_i^T\mathbf{l} = \widetilde{\mathbf{e}}_i^T\mathbf{U}_P\mathbf{a} = (\mathbf{U}_P^T\widetilde{\mathbf{e}}_i)^T\mathbf{a} = \bar{\mathbf{e}}_i^T\mathbf{a}$.

2.5. Conclusive remarks

In this Chapter we presented the parameterizations for both 2D and 3D ray spaces. The ray space is a parametric space of acoustic rays, i.e. each point in the ray space corresponds to a ray in the geometric space. The geometric primitives that compose the acoustic scene are represented in the ray space as sets of rays that intersect them. In the 2D ray space each geometric primitive is represented by a number of linear constraints. In the 3D ray space the geometric primitives are represented by a variable number of linear constraints and one quadratic constraint.

In the following Chapters the ray space representation of acoustic primitives will be used for both modeling and analysis purposes. As far as modeling is concerned, we will use the ray space representation of acoustic primitives, in particular that of the oriented reflector, to compute the mutual visibility between all reflectors in the environment. Successively this visibility information will be used for an efficient tracing of acoustic beams in the ray space.

For analysis purposes we take the opposite approach, i.e. instead of using the ray space representation of known acoustic primitives to model an unknown soundfield, we map the acquired acoustic measurements in the ray space and than we use this data to extract information about the unknown acoustic primitives. Working in the ray space, in which we known how the acoustic primitives should be represented, helps us detect, identify and extract the desired information.

Chapter 3

Modeling of sound propagation in the ray space

THIS Chapter describes the construction of the visibility data structure and a beam tracing method for modeling of acoustic propagation in complex 2D/3D environments, which exploits the information on mutual visibility between reflectors to speed up the computation of the acoustic beams. The visibility information depends on the environment geometry only and is thus evaluated in an offline phase. The use of the visibility data structure helps lighten the computational burden of the beam tracing phase as all occlusion handling is performed in the visibility precomputation stage. In practical terms this means no occlusion handling is needed during the beam tracing phase and, similarly to the path tracing phase that consists of checking if the acoustic path is a part of a given beam, the beam tracing consists of checking if the current beam is a part of the given visibility region.

3.1. Motivations

The acoustic room simulation finds applications in different fields. Some classical examples are: video-games and other virtual reality applications; building acoustics software; and music processing. However, the importance of predicting the

reverberations within an environment is non limited to just these areas. Recently a number of acoustic signal processing algorithms aimed at exploiting the knowledge of the environment appeared in literature [1]. Such methods use the propagation modeling engine to compute the early reflections and devise techniques to: model the soundfield of a virtual environment that has to be reproduced by a loudspeaker array [8], [9], [10]; compensate and exploit effects of reverberations in rendering applications [53]; use the energy of reflective paths to estimate the reflection coefficients of walls within the environment (see the Chapter 6); etc.

In fact, as the beam tracing method computes acoustic beams it turns out to be suitable not just for the binaural rendering, where the user experience is designed to produce the accurate impression of the acoustic scene from a pre-defined location in space, but also for the rendering of a soundfield in a region of space. The beam tracing geometrically computes the visibility of image sources from an entire region, which is the information that can be exploited by a rendering engine to enable an iterative rendering experience. In fact, interactivity of the rendering experience heavily relies on the possibility to spatially "explore" the acoustic scene by moving in it and, in doing so, we experience occlusions of virtual sources as a natural way to understand and interact with the environment. An extreme example of this is the ability to navigate in complex environments on the part of vision-impaired people.

3.2. Related work

As pointed out previously, the computational effort of beam tracing algorithm is unevenly divided between the beam tracing and paths tracing phases. Different solutions were proposed to address this problem and accelerate the construction of the beam tree. In [14] the Binary Space Partitioning (BSP) technique was used to divide the geometric space into convex regions. During the construction of the beam tree a beam is tested for incidence only on the reflectors of the current region and successively, if not completely occluded, the beam is propagated into the next region. This procedure avoids testing all possible reflectors at each iteration. The BSP technique is very popular for visibility preprocessing [54], [55] and was used not just for beam tracing but also for visibility tests in image-source methods [56]. However performance issues are still present for moving sources. In [31] the beam tree is constructed without accurate visibility calculations. Therefore the path tracing phase requires occlusion checks similar to those in the image source method, but the

construction of beams keeps the number of generated image sources under control. This acceleration technique allows modeling of moving sources in simple environments. Frustum tracing [57], [58] casts a finite number of small beams (frusta) from the source and propagates them through the environment similarly to rays in the ray tracing method. No accurate visibility computations are done - if needed the frusta are subdivided into smaller ones to minimize the error. This method is a combination of ray and beam tracing methods: it is generally much faster than beam tracing method and minimizes the aliasing problem of the ray tracing method. While suitable for real-time operation in generic 3D environments the frustum tracing suffers from inaccuracies in the modeling of acoustic paths.

In this work we are interested in gaining speed in the construction of the beam tree without sacrificing the accuracy of the beam tracing technique. Unlike [31] that postpones the visibility computation to the path tracing phase (the point-to-point visibility is more easily handled than the visibility from point) we anticipate the visibility computation in a phase that precedes the beam tracing. This obviously means handling a higher-order information and a more difficult problem to solve, but at this point it does not depend on source position and can be computed in advance. The BSP can be seen as a type of visibility precomputation as it determines the potentially visible set of reflectors for the beam in a given cell. However, this visibility information is not complete nor exact and visibility computations are still necessary during the beam tracing phase. In this work we are interested in computing a visibility information that releases the beam tracing phase from any occlusion handling and transforms it in just a series of intersection tests with regions of the new visibility data structure.

3.3. Problem formulation

At each iteration of the beam tracing procedure the beams are reflected with respect to reflectors they encounter first during their propagation. Therefore, the reflected bundle of rays has the origin (image source) behind the reflector that caused the reflection to occur and all its rays intersect the reflector of origin. Furthermore, the reflected bundle of rays is subdivided into sub-beams, each falling onto a different reflective surface. Consequently, each beam represents a subset of the reflector-to-reflector visibility region. Therefore, the visibility information sufficient for an efficient tracing of acoustic beams is the mutual visibility among all reflectors in the

environment. A general overview of the visibility-based beam tracing algorithm is shown in Figure 3.1 with illustrations depicted in a 2D case.

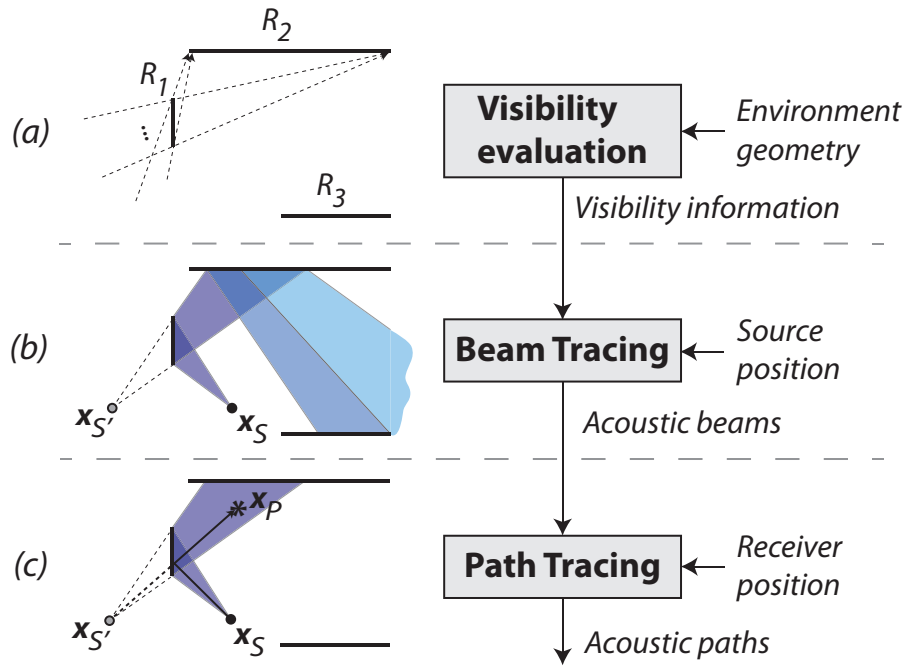


Figure 3.1.: The visibility-based beam tracing algorithm: (a) evaluation of mutual visibility between reflectors; (b) tracing of acoustic beams as they bounce and split during the interaction with the environment; (c) tracing of acoustic paths between source and receiver.

Visibility evaluation – The geometry of the environment is used for the evaluation of the mutual visibility between all reflectors in the environment. In particular, the visibility information is encoded using a suitable representation, which leads to the definition of an efficient visibility data structure describing all the rays reflected from a given wall. With reference to Figure 3.1-(a), the visibility of the reflector R_2 from the reflector R_1 is defined as the set of rays passing through both R_1 and R_2.

Beam Tracing – Given the source position $\boldsymbol{x}_S$, the beam reflected by R_1 can be thought as originated from the image source $\boldsymbol{x}_{S'}$ as shown in Figure 3.1-(b). The reflected beam is therefore a subset of rays going from R_1 to R_2 and can be computed efficiently by a lookup of the visibility data structure. This procedure continues iteratively as beams continue bouncing off the reflectors in the environment. During the interaction with the environment, beams split and branch. Some beams fall onto other reflectors, others continue to infinity as depicted in Figure 3.1-(b). The beams are organized in a beam tree data structure that contains the branching

relationship between acoustic beams and efficiently represents the visibility of the geometric space from the source position.

Path Tracing – Given the receiver position $\boldsymbol{x}_P$, the acoustic path between source and receiver can be seen as a ray joining the image source and the receiver (see Figure 3.1-(c)). Therefore, the path tracing is performed efficiently as a lookup of the beam tree, i.e. testing for the presence of the receiver inside the nodes (beams) of the beam tree data structure.

All the outputs of three main steps of the visibility-based algorithm are based on rays as the elementary primitive: mutual visibility between two reflectors is defined by the set of rays intersecting both reflectors without being occluded; beams are sets of rays departing from the same real/image source and falling to the same reflector; acoustic paths are rays joining the sources and the receiver. As a consequence, the use of the ray space for the representation and computation of visibility information and extraction of acoustic beams, turns out to be very useful.

We describe now the representation and organization of the visibility information, then we concentrate on the use of the described visibility data structure for an efficient tracing of acoustic beams. A detailed description of all the implementation issues can be found in literature, and it is not discussed here as it goes beyond the scope of the dissertation. For comprehensive surveys of visibility computation algorithms see [59], [60], [52]. For the 2D algorithm refer to the previous implementation in [32]. An algorithm for the visibility computation in the projective 2D ray space can also be found in [43]. Unlike the visibility evaluation for 2D environments, the visibility evaluation for generic 3D environments is not a linear problem anymore. Consequently this is a more challenging problem. Note however that the visibility evaluation is a pre-processing operation and therefore it is not time-critical. For visibility culling algorithms in Plucker space see the approaches described in [61] and [42] and further improved in [62], [63], [64], [65] and [66].

3.4. 2D Algorithm

In this section we present the 2D algorithm. While it has a minor impact compared to the 3D algorithm, we present it before as it is a necessary step for correct understanding of the 3D algorithm. Notice also that the 2D algorithm can easily be applied to a specific type of 3D environments defined as the Cartesian product of a 2D and a 1D environment (e.g. environments described by vertical walls ending in

perpendicular floor and ceiling) as described in [32].

3.4.1. Visibility information

The visibility information is used for speeding up the tracing of acoustic beams as they bounce around in the environment. The reflected beams are originated from an image (mirrored) source. In accordance with the image source principle, when we evaluate the visibility of the environment from a mirrored source, we do not consider the reflectors in the half-space where the mirrored source lies. Here the definition of oriented reflectors (2.6) turns to be useful.

Visibility region – All rays originated from a reflector R_{AB} (given by the line segment $\boldsymbol{x}_A\boldsymbol{x}_B$) falling onto reflector R_{CD} (given by the line segment $\boldsymbol{x}_C\boldsymbol{x}_D$) form the visibility region of R_{CD} from R_{AB}, indicated as $\mathcal{V}(\overline{R}_{AB}, \overline{R}_{CD})$. In the ray space this visibility region is given by the intersection of corresponding oriented reflectors (2.6), i.e.

$$\mathcal{V}(\overline{R}_{AB}, \overline{R}_{CD}) = \left(\overline{\mathbf{x}}_A^+ \cap \overline{\mathbf{x}}_B^-\right) \cap \left(\overline{\mathbf{x}}_C^+ \cap \overline{\mathbf{x}}_D^-\right).$$

As a consequence the visibility region is formed in the ray space by the intersection of four half-spaces as shown in Figure 3.2.

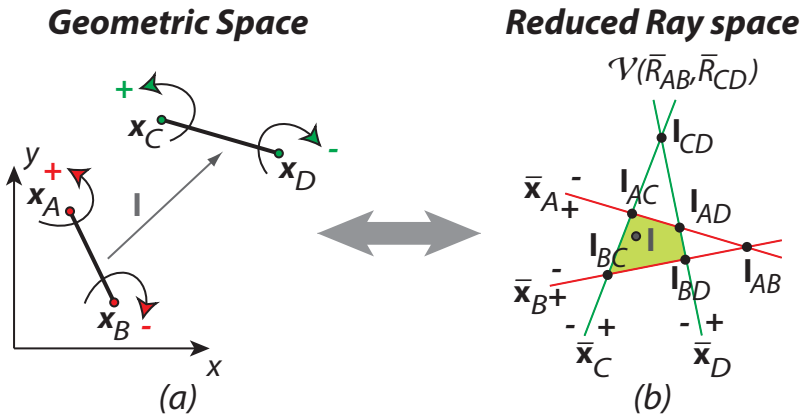


Figure 3.2.: (a) All rays originated from a reflector R_{AB} falling onto reflector R_{CD} form the visibility region of R_{CD} from R_{AB}; (b) The intersection of four half-spaces forms a visibility region in the ray space.

Visibility diagram – If the environment is composed by more than two reflectors, mutual occlusions may arise. This corresponds to an overlapping of visibility regions in the ray space. The resulting collection of visibility regions (overridden according to the front-to-back order of reflectors) constitutes the visibility diagram of the reflector

R_{AB}, $\mathcal{D}(\overline{R}_{AB}) = \{\mathcal{V}^*(\overline{R}_{AB}, \overline{R}_i) \neq \emptyset\}$, where R_i are the reflectors visible by the reflector R_{AB} and $*$ indicates that visibility regions have been overridden according to the front-to-back order (see Figure 3.3). As an example, given a reflector R_{EF} occluded by the reflector R_{CD}, we have

$$\mathcal{V}^*(\overline{R}_{AB}, \overline{R}_{EF}) = \mathcal{V}(\overline{R}_{AB}, \overline{R}_{EF}) - \left(\mathcal{V}(\overline{R}_{AB}, \overline{R}_{CD}) \cap \mathcal{V}(\overline{R}_{AB}, \overline{R}_{EF})\right).$$

Thus, the construction of a visibility diagram requires a subtraction of occluded

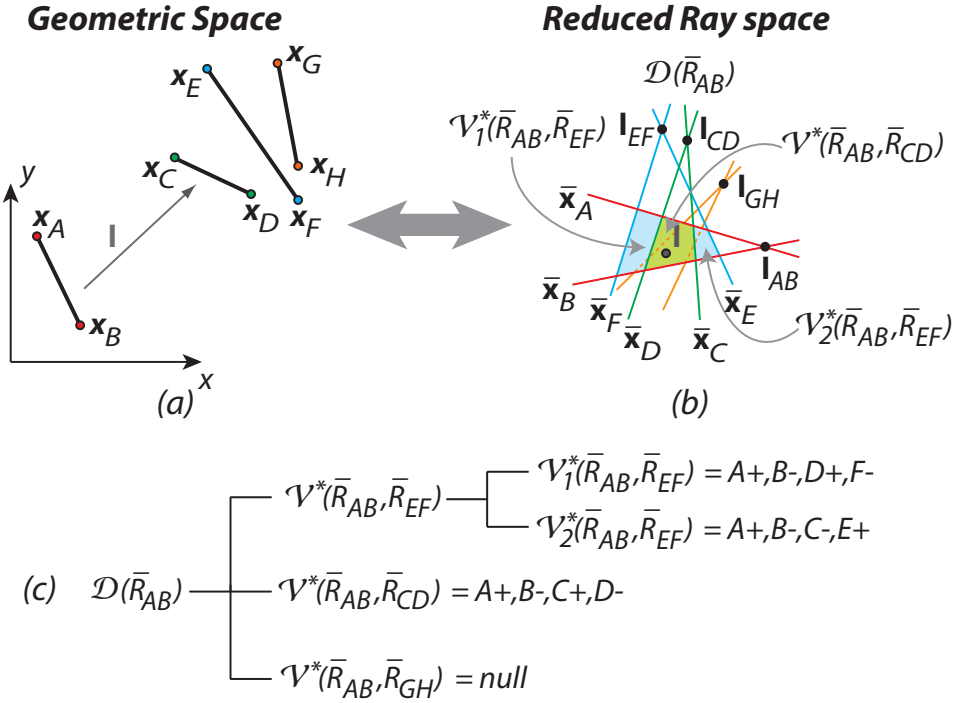


Figure 3.3.: The collection of visibility regions, overridden according to the front-to-back order of reflectors, constitutes the visibility diagram of the reflector R_{AB}.

portions of visibility regions, i.e. adding of new constraints, removal of redundant ones and possibly the splitting of a visibility region. An example of construction of the visibility diagram from individual visibility regions is shown in Figure 3.3 (b). The tree in Figure 3.3 (c) shows the decomposition of the visibility diagram into overridden visibility regions The construction of a visibility diagram depends only on environment geometry and thus it can be computed in an off-line phase, i.e. before the source or receiver positions are specified. Finally, the collection of visibility diagrams of all the reflectors gives us the information on the mutual

visibility between all reflectors in the environment.

At this point it is important to stress the fact that all the occlusion handling is performed at this stage. In the following beam tracing stage, beams are tested just for the intersection with the leaf nodes of the visibility diagram: interactions between reflectors do not need to be taken into account; the reflectors that are never visible from the reflector that originates the beam are not considered at all as they are already classified as invisible.

3.4.2. Beam tracing

The visibility conditions cause the splitting of rays originating from an image source into beams. With reference to Figure 3.4, we show how we use visibility diagrams, which encode the mutual visibility between reflectors, to iteratively trace beams. For each beam that falls onto a reflector we perform the following steps:

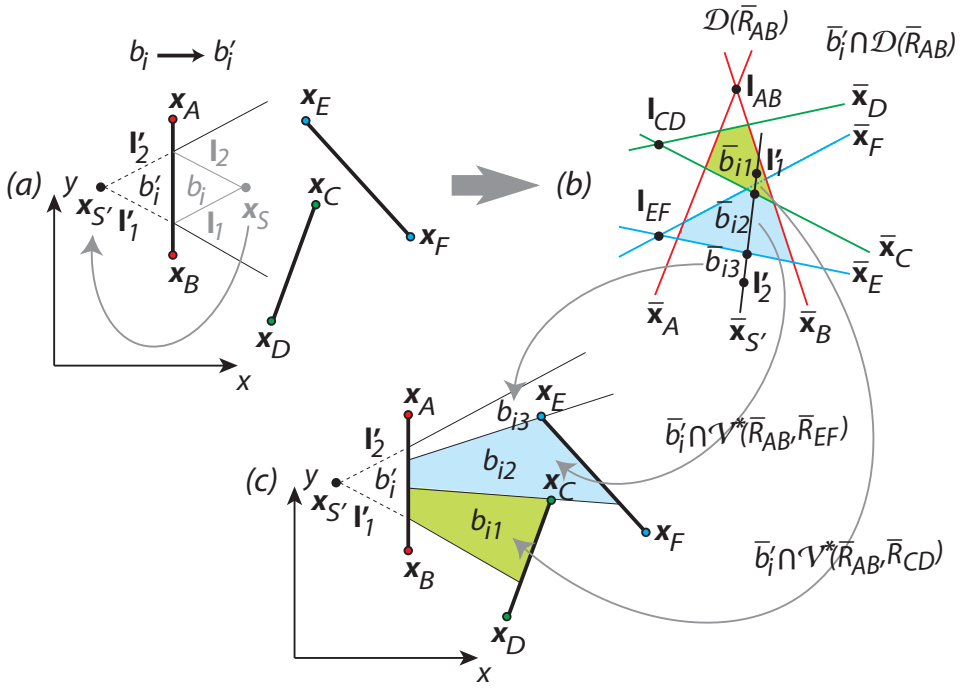


Figure 3.4.: (a) Reflection of the beam in the geometric space; (b) Beam subdivision in the ray space; (c) New beams in the geometric space.

Step 1: *Compute the reflected bundle of rays* – let us consider, for example, the reflection of a beam b_i onto the reflector R_{AB}, shown in Figure 3.4 (a). We first

compute the reflected bundle of rays b_i'. This is done mirroring the beam's source $\boldsymbol{x}_S$ and the two edge rays that limit the beam, $\mathbf{l}_1$ and $\mathbf{l}_2$, with respect to the reflector line $\mathbf{l}_{AB}$ using the reflection matrix $\mathbf{M}^{AB}$ [67]

$$\begin{aligned}
&\mathbf{x}_{S'} = \mathbf{M}^{AB}\mathbf{x}_S, \quad \mathbf{l}' = \det(\mathbf{M}^{AB})(\mathbf{M}^{AB})^{-T}\mathbf{l} \\
&\mathbf{M}^{AB} = \mathbf{I}_3 - 2\mathbf{G}\frac{\mathbf{l}_{AB}\mathbf{l}_{AB}^T}{\mathbf{l}_{AB}^T\mathbf{G}\mathbf{l}_{AB}}, \ \mathbf{G} = \begin{bmatrix} \mathbf{I}_2 & 0 \\ 0 & 0 \end{bmatrix}, \\
&\mathbf{l}_{AB} = \mathbf{x}_A \times \mathbf{x}_B,
\end{aligned}$$

where $-T$ indicates the the transpose of the inverse [1];

Step 2: *Split the reflected bundle of rays into sub-beams* – after obtaining the reflected bundle of rays b_i' the splitting process is accomplished in the ray space by intersecting the reflector's visibility diagram, $\mathcal{D}(\overline{R}_{AB})$, with the ray space representation of b_i', $\overline{b}_i'$, as shown in Figure 3.4(b). The $\overline{b}_i'$ is made of a number of segments, each lying in a different visibility region. These segments represent the sub-beams originated from the splitting of b_i'. The corresponding beams in the geometric space are depicted in Figure 3.4(c). Some of them proceed to infinity, others are blocked by reflectors and therefore they originate new beams;

Step 3: *Add new beams to the beam tree data structure* – the beam b_i represents a node of the beam tree; it is reflected from the reflector and the reflected bundle b_i' is split into sub-beams that represent the new branches of the beam tree departing from the node b_i;

Step 4: *Repeat the procedure for each beam that falls onto a reflector* – the recursive procedure stops when the preassigned order of reflection is reached or when the beams die out (i.e., when they are attenuated below a preassigned threshold of magnitude).

3.4.3. Path tracing

Beams are organized in a beam tree data structure that contains the branching relationship between acoustic beams and represent efficiently the visibility from the geometric space of the source position. Once the receiver location is specified, a simple iterative procedure looks up the beam tree to find the paths from source to

[1] The ray $\mathbf{l}'$ can be seen as passing through two points $\boldsymbol{x}_{P_1}$ and $\boldsymbol{x}_{P_2}$; using the relation (2.3) and the propriety of the cross product $(\mathbf{M}^{AB}\mathbf{x}_{P_1}) \times (\mathbf{M}^{AB}\mathbf{x}_{P_2}) = \det(\mathbf{M}^{AB})(\mathbf{M}^{AB})^{-T}(\mathbf{x}_{P_1} \times \mathbf{x}_{P_2})$, we obtain the given relation between $\mathbf{l}'$ and $\mathbf{l}$.

receiver. Each node in the beam tree corresponds to a beam and a simple geometrical test checks if a receiver is inside the beam. A number of different implementations of the test can be devised. Two possibilities are shown in Figure 3.5. In Figure 3.5 (a) the beam is parametrized with the four bounding lines oriented so that a point inside the beam is always on the right of those lines. In order to test if a point is inside the beam, therefore, we only need to verify that it is on the right side of all lines that parametrize the beam, i.e. compute four scalar products. In Figure 3.5 (b) the beam is parametrized using two reflector lines and the two points that limit the beam. The presence of the receiver is tested checking if the receiver is between the two reflectors (if the receiver is on the correct side of each reflector) and if the path going from the beam's source to the receiver is inside the beam (if it has the correct orientation with respect to the points that limit the beam).

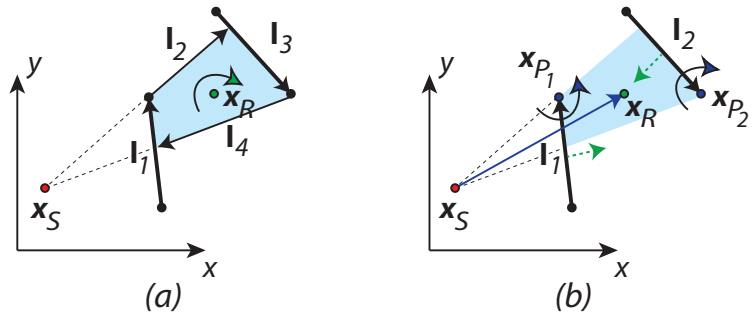


Figure 3.5.: Two possible parametrizations of the beam: (a) the beam is parametrized with the four bounding lines oriented so that a point inside the beam is always on the right of those lines; (b)the beam is parametrized using two reflector lines and the two points that limit the beam.

3.5. 3D Algorithm

3.5.1. Visibility information

The definition of visibility information in 3D environments follows the ideas explained for 2D geometry. However, the higher dimensional space and a necessary intersection with a non-convex quadric surface makes the implementation more complex.

We define visibility region between two reflectors $\mathcal{V}(\overline{R}_i, \overline{R}_j)$ (intersection of all rays that go from one reflector and fall onto the other one) and visibility diagram of a reflector $\mathcal{D}(\overline{R}_i)$ (the collection of visibility regions from that reflector) as in the

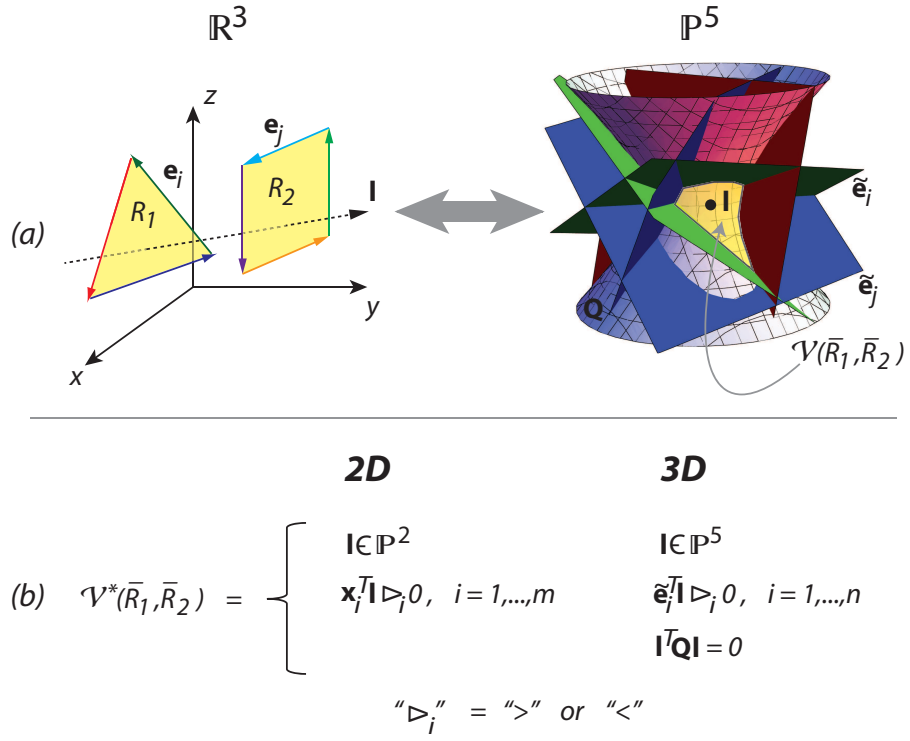


Figure 3.6.: (a) Visibility region in 3D geometric and ray spaces; (b) Visibility region constraints for 2D and 3D geometries.

2D case. Similarly to the 2D case the visibility region is given by the intersection of corresponding oriented reflectors (2.11) as shown in Figure 3.6 (a). The visibility diagram is the collection of overlapped visibility regions. As in 2D, after overlapping the visibility regions are given by the intersection of a number of half-spaces (this ensures that rays have the correct orientation). However there is also the intersection with the Plucker quadric (this ensures that the rays that define the region are real lines in 3D). A comparison between the visibility regions in 2D and 3D is shown in Figure 3.6 (b).

While evaluating the mutual visibility between reflectors in the environment only events (intersections) on the surface of Plucker quadric have to be taken into account, i.e. only events that correspond to different visibility changes in 3D. As a consequence, the removal of occluded portions of visibility regions is no more a linear program, and a new, dedicated, algorithm for detecting occlusions, adding/removing constraints and splitting regions has to be developed. However, as already stated, presenting this algorithm goes beyond the scope of this work. A simple pseudocode

for the visibility evaluation algorithms should look like:

function Visibility evaluation ($\overline{R}_i$, $\forall i$)
begin
  for all reflectors $\overline{R}_i$ **do**
    initialize the visibility diagram $\mathcal{D}(\overline{R}_i) = \emptyset$
    for all reflectors $\overline{R}_j$, $j \neq i$ **do**
      initialize $\mathcal{V}^*(\overline{R}_i, \overline{R}_j) = \emptyset$
      compute the visibility region $\mathcal{V}(\overline{R}_i, \overline{R}_j)$
      add $\mathcal{V}(\overline{R}_i, \overline{R}_j)$ to $\mathcal{V}^*(\overline{R}_i, \overline{R}_j)$
      for all possible occluders[a] $\overline{R}_k$ **do**
        compute the visibility region $\mathcal{V}(\overline{R}_i, \overline{R}_k)$
        for all $\mathcal{V}_r^*(\overline{R}_i, \overline{R}_j) \in \mathcal{V}^*(\overline{R}_i, \overline{R}_j)$ **do**
          test for intersection $\mathcal{V}_r^*(\overline{R}_i, \overline{R}_j) \cap \mathcal{V}(\overline{R}_i, \overline{R}_k)$
          if intersection = true **then**
            subdivide $\mathcal{V}_r^*(\overline{R}_i, \overline{R}_j)$ into $\mathcal{V}_o^*(\overline{R}_i, \overline{R}_j)$[b]
            remove $\mathcal{V}_r^*(\overline{R}_i, \overline{R}_j)$ from $\mathcal{V}^*(\overline{R}_i, \overline{R}_j)$
            add $\mathcal{V}_o^*(\overline{R}_i, \overline{R}_j)$ to $\mathcal{V}^*(\overline{R}_i, \overline{R}_j)$
          end
        end
      end
      add the overridden regions $\mathcal{V}^*(\overline{R}_i, \overline{R}_j)$ to $\mathcal{D}(\overline{R}_i)$
    end
  end
  return $\mathcal{D}(\overline{R}_i)$, $\forall i$
end

[a] the occluders can be sorted by importance by shooting test rays and measuring, for each occluder, the number of rays intersecting it, as suggested in [62] and [63];
[b] empty set in case of the complete occlusion.

The reader should consult the listed references for more details.

3.5.2. Beam tracing

The procedure is similar to the 2D case. With reference to Figure 3.7 (a) consider the reflection of the beam b_i, with source $\boldsymbol{x}_S$ and edge constraints $\widetilde{\mathbf{e}}_i$, $i = 1, ..., n$, from the reflector R_1.

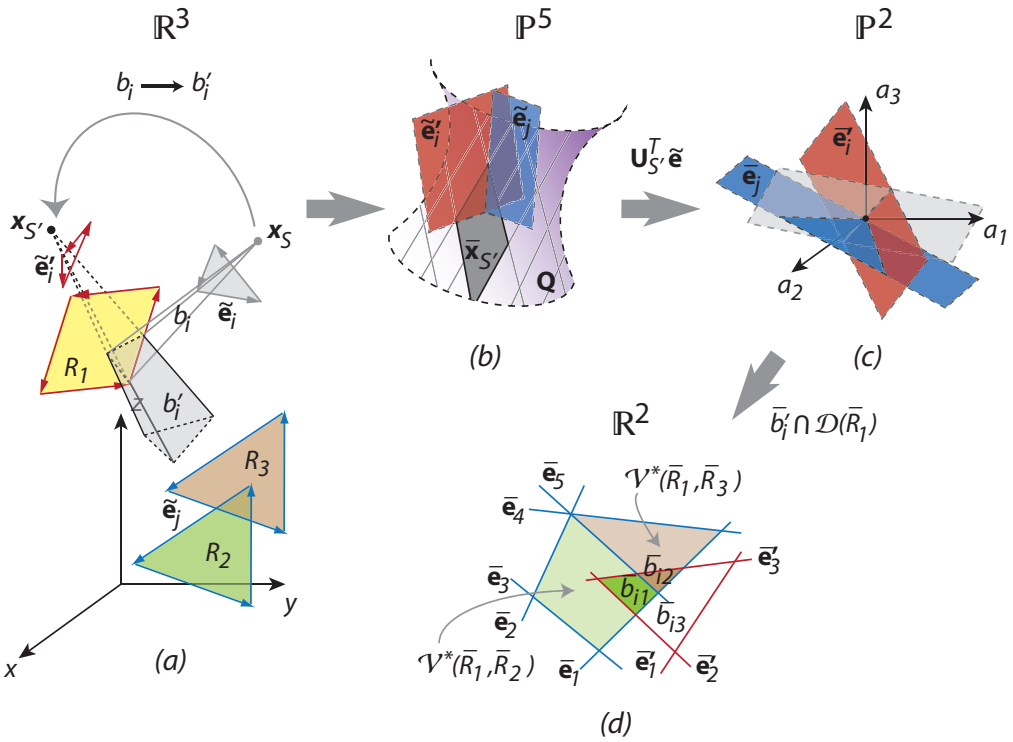


Figure 3.7.: (a) Reflection of a beam from the reflector; (b) The reflected beam in the ray space; (c) Subspace of rays originated by the image source $\boldsymbol{x}_{S'}$; (d) Subdivision into sub-beams.

Step 1: First the source $\boldsymbol{x}_S$ and edges $\tilde{\mathbf{e}}_i$ are reflected with respect to the plane on which the reflector lies, $\pi = [A, B, C, D]^T$. The result is the reflected bundle of rays b_i' with image source $\boldsymbol{x}_{S'}$ and constraints $\tilde{\mathbf{e}}_i'$, $i = 1, ..., n$. The reflection is performed using the reflection matrix

$$\mathbf{M}^\pi = \mathbf{I}_4 - 2\mathbf{G}\frac{\pi\pi^T}{\pi^T\mathbf{G}\pi} = \begin{bmatrix} \mathbf{N} & \mathbf{n} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{I}_3 & 0 \\ 0 & 0 \end{bmatrix}.$$

Using the homogeneous coordinates we have $\mathbf{x}_{S'} = \mathbf{M}^\pi \mathbf{x}_S$ and therefore the Cartesian coordinates of the image source $\boldsymbol{x}_{S'}$ are obtained as $\boldsymbol{x}_{S'} = \mathbf{N}\boldsymbol{x}_S + \mathbf{n}$. Given a generic edge passing through two points $\boldsymbol{x}_{P_1}$ and $\boldsymbol{x}_{P_2}$

$$\tilde{\mathbf{e}} = \begin{bmatrix} \boldsymbol{x}_{P_1} \times \boldsymbol{x}_{P_2} \\ \boldsymbol{x}_{P_2} - \boldsymbol{x}_{P_1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}^m \\ \mathbf{e}^d \end{bmatrix},$$

it is reflected with respect to π as follows

$$\begin{aligned}\widetilde{\mathbf{e}}' &= \begin{bmatrix} (\mathbf{N}\boldsymbol{x}_{P_1}+\mathbf{n})\times(\mathbf{N}\boldsymbol{x}_{P_2}+\mathbf{n}) \\ (\mathbf{N}\boldsymbol{x}_{P_2}+\mathbf{n})-(\mathbf{N}\boldsymbol{x}_{P_1}+\mathbf{n}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{N}\boldsymbol{x}_{P_1}\times\mathbf{N}\boldsymbol{x}_{P_2}+\mathbf{n}\times\mathbf{N}(\boldsymbol{x}_{P_2}-\boldsymbol{x}_{P_1})+\mathbf{n}\times\mathbf{n} \\ \mathbf{N}(\boldsymbol{x}_{P_2}-\boldsymbol{x}_{P_1}) \end{bmatrix} \\ &= \begin{bmatrix} \det(\mathbf{N})(\mathbf{N}^{-1})^T\mathbf{e}^m+\mathbf{n}\times(\mathbf{N}\mathbf{e}^d)+\mathbf{n}\times\mathbf{n} \\ \mathbf{N}\mathbf{e}^d \end{bmatrix}.\end{aligned}$$

In the ray space the reflected bundle of rays b'_i, $\overline{b}'_i$, lies completely on the surface of Plucker quadric and is limited by the hyperplanes $\widetilde{\mathbf{e}}'_i$, $i = 1, ..., n$, as illustrated in Figure 3.7 (b).

Step 2: The visibility conditions, encoded as constraints on edges $\widetilde{\mathbf{e}}_j$, $i = 1, ..., m$, cause the subdivision of b'_i into sub-beams. At this point, to perform this subdivision there is no need to remain in a six-dimensional space and deal with a non-convex quadric surface. In fact, the reflected bundle of rays b'_i has the origin in the image source $\boldsymbol{x}_{S'}$. The ray space representation of $\boldsymbol{x}_{S'}$, $\overline{\mathbf{x}}_{S'}$, is the projective plane (2.14) spanned by (2.12) that lies completely on the surface of the Plucker quadric. As a consequence $\overline{b}'_i$ lies on the surface of the Plucker quadric as well, and it is represented by the portion of the linear subspace $\overline{\mathbf{x}}_{S'}$. Using (2.15) we can project both the beam constraints $\widetilde{\mathbf{e}}'_i$ and the visibility constraints $\widetilde{\mathbf{e}}_j$ to the subspace $\overline{\mathbf{x}}_{S'}$ that defines only the rays passing through the image source $\boldsymbol{x}_{S'}$. The result is a collection of linear constraints, $\bar{\mathbf{e}}'_i$ and $\bar{\mathbf{e}}_j$, within a two-dimensional projective space $\mathbb{P}^2$ as depicted in Figure 3.7 (c).

Step 3: The reflected bundle of rays b'_i is then subdivided into new beams intersecting the reflector's visibility diagram $\mathcal{D}(\overline{R}_1)$, i.e. $\overline{b}'_i \cap \mathcal{D}(\overline{R}_1)$. This procedure is performed in the space spanned by the homogeneous coordinates $\mathbf{a}$. As done previously for the 2D reduced ray space, for clarity of visualization we depict the splitting process in a two-dimensional Euclidean space obtained intersecting $\mathbb{P}^2$ with a prescribed plane. A simple example of the visibility diagram $\mathcal{D}(\overline{R}_1)$ with two computed visibility regions, $\mathcal{V}^*(\overline{R}_1, \overline{R}_2)$ and $\mathcal{V}^*(\overline{R}_1, \overline{R}_3)$, is shown in Figure 3.7 (d). The overlapping between the regions has been managed during the visibility precomputation stage and the visibility diagram shows not occluded convex portions defined by a number of linear constraints $\bar{\mathbf{e}}_j$. Finally, the new beams are computed intersecting the reflected bundle of rays b'_i, specified with the constraints $\bar{\mathbf{e}}'_i$, with the regions

of the visibility diagram $\mathcal{D}(R_1)$ as shown in Figure 3.7 (d). Some beams encounter reflectors along their path (b_{i1} and b_{i2}) and originate new reflections, others proceed to infinity (b_{i3}).

3.5.3. Path tracing

In order to test the presence of the receiver inside the beam a parametrization similar to the parametrization used for 2D and shown in Figure 3.5 (b) is used. The beams are parametrized using two planes that contain the reflectors and a number of edges that limit the beam. The presence of the receiver is tested checking if the receiver is between the two planes and if the path going from the beam's source to the receiver is inside the beam. Examples of path tracing in environments with different complexity are shown in Figure 3.8.

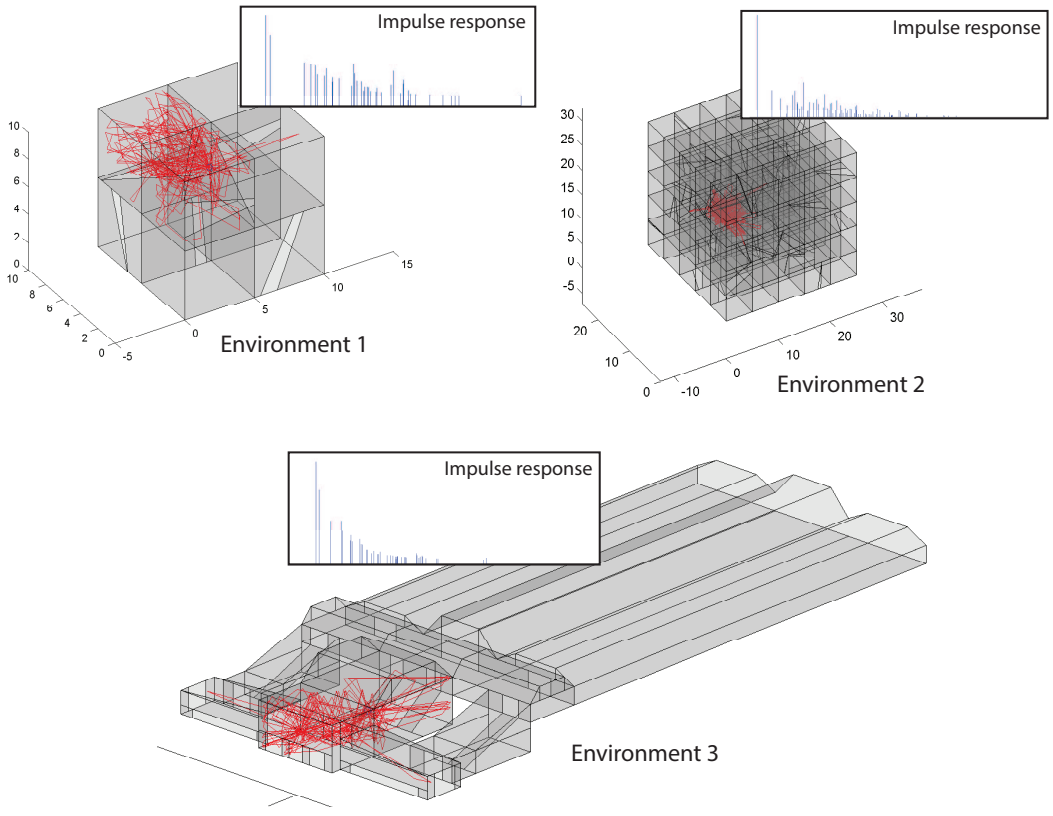


Figure 3.8.: Examples of path tracing in environments with different complexity.

3.5.4. Considerations

As seen, the increase in complexity of problems encountered while generalizing the 2D visibility approach to the 3D case is not proportional to the increase in the dimension of the geometric space. The visual events become, in general, non-linear and non-convex. This is reflected on the definition of the ray space represented, in 3D case, to be a four-dimensional non-convex quadric surface embedded in $\mathbb{P}^5$. It is important, however, to note that the increase in dimension of the beam tracing phase is exactly one. Observing Figures 3.4 and 3.7 we note that: all rays passing through a point in 2D have 1 degree-of-freedom, i.e. in the ray space they belong to a projective line $\mathbb{P}^1$; in 3D the rays passing through a point have 2 degrees-of-freedom, i.e. in the ray space they belong to a projective plane $\mathbb{P}^2$. Furthermore, in 3D the subspace of all rays passing through the point lies completely on the surface of the ruled quadric, thus no intersection is needed. Indeed, the visibility events from a point are, once again, linear as observed in [50], [46], [68], [69], [59], [42]. The beam tracing is thus a linear program that consist of the intersection of the reflected beam with the precomputed visibility regions of the source reflector. As a consequence, even in 3D, the beam tracing procedure has a reasonable computational burden that allows efficient tracing of acoustic beams.

The visibility evaluation requires only the information on environment geometry and is computed off-line. The resulting visibility diagrams do not require much memory space, as they are represented just by a list of active constraints (borders) of each visibility region. Therefore the visibility information is easily stored and loaded together with the environment. In general, the dynamic environments are not easily handled as they require recomputing the visibility information. However, some simple dynamic scenes that require just a removal and/or re-introduction of some reflectors (e.g. opening and closing of doors and windows) can be handled by simply combining the visibility regions of the removed reflector with the visibility regions of reflectors that observe it.

3.6. Conclusive remarks

In this Chapter we presented the 2D and 3D visibility based beam tracing algorithms. Both 2D and 3D algorithms make use of the visibility information computed in the 2D and 3D ray spaces, respectively. Visibility information specifies the mutual

visibility between all the reflectors in the environment. It is made by a collection of visibility regions that describe the visibility conditions between two reflectors using the ray space representation of corresponding oriented reflectors. In 2D, the visibility region is defined by a number of linear constraints. In 3D, the visibility region is defined by a number of linear constraints and one quadratic constraint. The beam tracing is performed subdividing the reflected bundle of rays into sub-beams by intersecting the ray space representation of the reflected beam with the visibility region of the reflector that causes the reflection to occur. In both 2D and 3D the beam tracing is a linear procedure.

The modeling engine is used in Chapter 6 for the computation of reflective paths. The modeled paths are then matched with the measurements in order to estimate the reflection coefficients of all reflective surfaces in the environment. For the rendering application in Chapter 7 the modeling engine is used to simulate the soundfield of both real (host) and virtual (to be reproduced) environments. This models are then used by the rendering engine to reproduce the acoustics of the virtual environment while compensating for the undesired reverberation of the real environment.

Chapter 4

Analysis of wavefields through plenacoustic imaging

ACOUSTIC scene analysis is generally performed by gathering measurements and combining the related constraints. This is done through a process that must be specifically developed for the problem at hand. In this Chapter we propose a methodology that allows us to follow a different route, which consists of collecting at once all the information that is available on the acoustic scene; organizing it into a data structure that displays it in a ready-to-interpret fashion; and performing the analysis of the collected data afterwards, using various methodologies, typically from pattern analysis.

The novel approach to acoustic scene analysis is based on near field acoustic measurements acquired with a microphone array and, in particular, on the concept of the plenacoustic function. We start from a general definition of the plenacoustic function, which includes directional information. We then show how to map this function onto the 2D ray space. By representing the plenacoustic (spatially distributed) measurements in the ray space, i.e. associating magnitudes to acoustic rays, we obtain snapshots (for a given time and frequency) of the plenacoustic function. This parameterization of the plenacoustic function is here referred to as a plenacoustic image. This image will turn out to carry a great deal of information on the acoustic

scene. We then show how to acquire an approximation of the above ideal plenacoustic image using a microphone array and describe how real and ideal images are related to each other. In particular, we discuss the impact of spatial sampling onto the resolution of the plenacoustic image (aperture problem), and the issue of scene visibility and its relation with the extension of the imaging array.

4.1. Motivations

Acoustic systems could greatly benefit from spatially distributed observations of the acoustic scene. In the near future we can expect to have a number of small microphone arrays distributed inside the environment (e.g. think about the array integrated in the Microsoft Kinect® device). Having at the disposal different "views" of the scene can help improve the performance and robustness of acoustic signal processing algorithms, for example merging different estimates to improve estimation accuracy or selecting the best view for the particular application; and accomplish tasks that would not be possible otherwise, such as taking into account or estimating visibility discontinuities of the environment. To address the problem of the representation and use of such a spatial data, we consider different views as snapshots of the plenacoustic function.

4.2. Related work

The plenacoustic function was introduced in [15] as the acoustic counterpart of the plenoptic function [33], which describes the intensity of the light flow at every position for every direction at all frequencies and, for dynamic scenes, time. In defining a plenoptic function, several assumptions are often made (e.g. static scenes, grayscale images, reduction of degrees of freedom on camera locations) in order to reduce the dimensionality of the representation [70]. Popular parametrizations are the lumigraph [71] and the lightfield [72], whose domain is the *space of oriented lines*. The plenoptic data is used in computer vision for localization, mapping, synthetic view generation (image based rendering), etc. An example of a commercial plenoptic camera is based on work in [73].

The plenacoustic function was originally proposed in [15] as the acoustic pressure as a function of space, time and frequency. In this definition the directional

information was dropped by the authors because of the lack of devices that could measure sound pressure and components of its gradient. The method was proposed and developed for reconstructing the sound pressure distribution anywhere from a measurements on a finite number of points. In this thesis, we revise this definition by re-introducing direction, thus making it completely symmetrical to its optical counterpart, as the directional information can be recovered with a microphone array through space-time processing. Redefining the plenacoustic function this way also induces a re-interpretation of the term "soundfield" as a vector field, which matches the concept of "lightfield" discussed in [72].

4.3. The plenacoustic camera

In this section we revise the definition of plenacoustic function by introducing the directional information, knowing that this information can be recovered through space-time processing. In fact, one way to measure the soundfield coming from a given direction in a given point in space is to use a beamforming algorithm with a compact microphone array. This way we can measure the magnitude (or power) of the acoustic pressure along rays that pass through its center. In other words, the compact array is used as an "acoustic camera", and the center of the array plays the same role as the optical center in a photo camera (center of collection of all rays). In a 2D (planar) geometry, such measurements are represented as a function of just one parameter, which is the direction if arrival of the acoustic ray. In this thesis we introduce a direct generalization of this idea by defining the "plenacoustic (or soundfield) camera", similarly to its optical counterpart, the plenoptic (lightfield) camera [74].

We define the soundfield camera as a device that acquires the plenacoustic function over a spatially extended "observation window" facing the acoustic scene. In the 3D case the observation window is a compact region of a planar surface, while in the 2D case it becomes a simple segment. If the observation window were infinitely extended (a plane or a line), then knowing the plenacoustic function on it would allow us to to infer it everywhere in space. This is indeed true because of Huygens Principle, but it is also true because of the principle of invariance of the radiance along acoustic rays. By limiting our knowledge of the plenacoustic function to a window of finite extension, we approximate this idea. The information that we gather through a finite window, in fact, only allows us to reconstruct the wave field in a region surrounding

the it, whose size depends on visibility and occlusion conditions.

In this work we define the soundfield camera in planar geometry and, therefore, we capture the plenacoustic function along a segment of a line, which represents the window of observation of the acoustic scene. This means that the soundfield camera captures the acoustic field along all possible rays that cross that segment from one given side. The plenacoustic function, measured along the considered segment and mapped onto the ray space will be referred to as *plenacoustic image* or *soundfield image*. Using the ray space representation of geometric primitives described in the Chapter 2 we can understand how an acoustic scene will appear in a plenacoustic image. The ray space is the projective space whose (homogeneous) coordinates are the parameters l_1, l_2 and l_3 of the line. We have seen that defining the ray space on the projective domain $\mathbb{P}^2$ (which uses three homogeneous coordinates) instead of the Euclidian domain $\mathbb{R}^2$ (which uses two coordinates) has several advantages. First of all, this choice is of general validity, as it accommodates rays coming from all possible directions (e.g. not necessarily just those coming from the front side). This means that this representation is independent of how we plan to implement the plenacoustic camera. In addition, in this space we can work with near-field sources or with sources at infinity (planar wavefronts) without having to change the notation; projective transformation such as specular reflections (wall-reflected rays or beams) are readily described with linear transformations (homographies). However, in order to simplify the visualization of plenacoustic images, we will often resort to showing a "reduced" ray space, whose domain in $\mathbb{R}^2$ is obtained by slicing the ray space (l_1, l_2, l_3) with a plane. In particular, unless otherwise stated, the cutting plane will be chosen as $l_1 = 1$, as often done in the area of computer vision. This particular choice of reduced ray space is fully equivalent to the projective ray space except for the rays that are parallel to the x-axis, which we cannot represent because they have $l_1 = 0$.

4.3.1. The ideal plenacoustic camera

An ideal plenacoustic camera is a device that is able to capture the (complex) amplitudes of all acoustic rays that fall onto it. The term "plenacoustic" comes from the fact that there is no single point (camera center) that rays are bound to pass through. A camera of this sort can be thought of as a line segment in the geometric space or, conversely, as the set of all rays that cross such segment.

Its representation in the ray space is therefore given by the parameters of all such rays. As a consequence its representation in the ray space is equivalent to that of a co-located reflector (2.7).

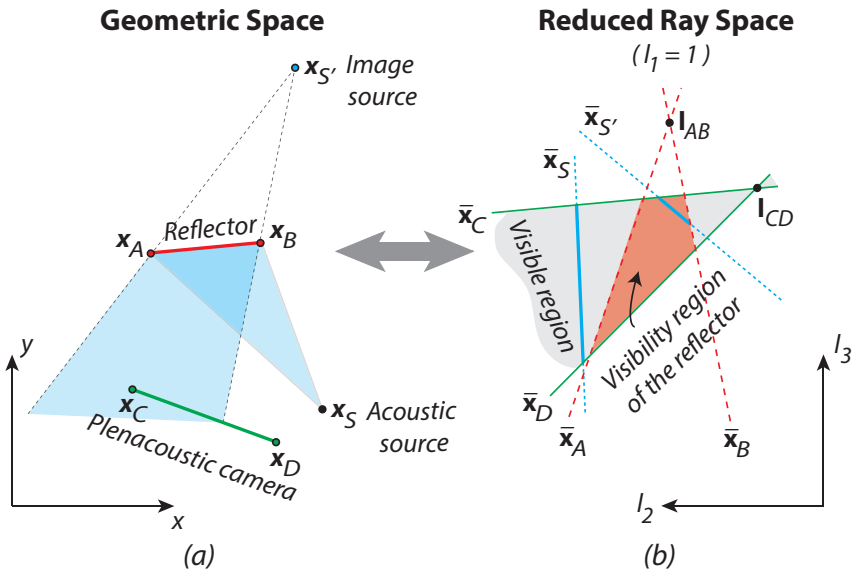


Figure 4.1.: (a) The ideal plenacoustic camera in the geometric space; (b) The ray space representation of the plenacoustic image.

So far we have discussed the representation of geometric primitives as a function of the acoustic rays that are generated (source), collected (receiver) or reflected (reflector) by them. Each ray that parameterizes the geometric primitive is associated to a magnitude. As soon as the acoustic source "lights up", a proliferation of acoustic rays populates the environment due to the presence of reflectors. Consider the simple scenario in Figure 4.1. The acoustic source $\boldsymbol{x}_S$ is visible from the plenacoustic camera $\boldsymbol{x}_C\boldsymbol{x}_D$. The reflector $\boldsymbol{x}_A\boldsymbol{x}_B$ causes the mirror source $\boldsymbol{x}_{S'}$ to appear. Notice, however, that reflective rays starting from $\boldsymbol{x}_{S'}$ are bound to pass through $\boldsymbol{x}_A\boldsymbol{x}_B$, thus limiting the portion of space where $\boldsymbol{x}_{S'}$ is visible. The bundle of reflective rays partially illuminates the acoustic camera $\boldsymbol{x}_C\boldsymbol{x}_D$. The corresponding situation in the ray space is depicted in Figure 4.1(b). In particular, the portion of ray space visible by the camera $\boldsymbol{x}_C\boldsymbol{x}_D$ is depicted in light gray, while the red area is the intersection between the regions of visibility of the reflector $\boldsymbol{x}_A\boldsymbol{x}_B$ and of the camera $\boldsymbol{x}_C\boldsymbol{x}_D$. Finally, when the source $\boldsymbol{x}_S$ is active, two segments in the visible region of $\boldsymbol{x}_C\boldsymbol{x}_D$ will "light up". The first is the intersection between $\overline{\mathbf{x}}_S$ (representation of $\boldsymbol{x}_S$ in

the ray space (2.4)) and the visible region, while the second one is the intersection among $\overline{\mathbf{x}}_{S'}$ ($\boldsymbol{x}_{S'}$); the visibility region of $\boldsymbol{x}_A\boldsymbol{x}_B$; and the region visible by $\boldsymbol{x}_C\boldsymbol{x}_D$.

Let $\mathcal{F}(\theta, x, y, t, f)$ be the plenacoustic function that specifies the acoustic wavefield coming from a given direction θ at a given position x, y at a given time t and frequency f. The ideal plenacoustic camera captures all the acoustic rays $\mathbf{l}(\theta, x, y)$ that impinge from all directions θ on the points x, y of the corresponding observation window $\boldsymbol{x}_C\boldsymbol{x}_D$, i.e. $\mathbf{l}(\theta, x, y) \in \overline{R}_{CD}$, where $\overline{R}_{CD}$ is the set of rays that form the ray space representation of the given line segment (2.7). Therefore the ideal plenacoustic image $\mathcal{I}^*(\mathbf{l})$ represents a snapshot of the plenacoustic function at time t_0, frequency f_0 and for directions and positions such that $\mathbf{l}(\theta, x, y) \in \overline{R}_{CD}$. The resulting image carries information on both magnitude and phase of the acoustic radiance, therefore it is generally complex-valued. If the application does not require phase information (as in the case of this dissertation), the image formation process simplifies. In this case it is convenient to construct the power plenacoustic image, i.e.

$$\mathcal{I}^*(\mathbf{l}) = \{|\mathcal{F}(\theta, x, y, t, f)|^2 \mid \mathbf{l} = \mathbf{l}(\theta, x, y) \in \overline{R}_{CD}, t = t_0, f = f_0\}. \tag{4.1}$$

Note however that depending on the application the phase information can be used as well.

The soundfield produced by the acoustic source $\boldsymbol{x}_S$ is mapped over the line $\overline{\mathbf{x}}_S$ in the ray space. Therefore the ray space image of $\boldsymbol{x}_S$ is

$$\mathcal{I}^*_S(\mathbf{l}) = a_S(\mathbf{l})\delta_S(\mathbf{l}), \tag{4.2}$$

where $a_S(\mathbf{l})$ is a function that depends on radiometric properties of the source, i.e. the signal amplitude and the radiation pattern for the direction $\mathbf{l}$ (at time t_0 and frequency f_0), and

$$\delta_S(\mathbf{l}) = \begin{cases} 1 & \text{if } \mathbf{l} \in \overline{\mathbf{x}}_S, \\ 0 & \text{if } \mathbf{l} \notin \overline{\mathbf{x}}_S. \end{cases}$$

On the other hand, the reflections produced by the image source $\boldsymbol{x}_{S'}$ are bound to pass through $\boldsymbol{x}_A\boldsymbol{x}_B$. Therefore the ray space image of $\boldsymbol{x}_{S'}$ is

$$\mathcal{I}^*_{S'}(\mathbf{l}) = a_{S'}(\mathbf{l})\delta_{S'}(\mathbf{l})r_{AB}(\mathbf{l})\Pi_{AB}(\mathbf{l}),$$

where $a_{S'}$ and $\delta_{S'}(\mathbf{l})$ are defined similarly to (4.2); $r_{AB}(\mathbf{l})$ is a function that depends

the reflective properties of the reflector, i.e. the reflection coefficient for the direction $\mathbf{l}$ (at frequency f_0), and

$$\Pi_{AB}(\mathbf{l}) = \begin{cases} 1 & \text{if } \mathbf{l} \in \overline{R}_{AB}, \\ 0 & \text{if } \mathbf{l} \notin \overline{R}_{AB}. \end{cases}$$

Finally, the plenacoustic image (4.1) contains the rays produced by both direct and image sources, $\boldsymbol{x}_S$ and $\boldsymbol{x}_{S'}$, that intersect the plenacoustic camera $\boldsymbol{x}_C\boldsymbol{x}_D$ (see the Figure 4.1). Therefore we have

$$\begin{aligned} \mathcal{I}^*(\mathbf{l}) &= (\mathcal{I}^*_S(\mathbf{l}) + \mathcal{I}^*_{S'}(\mathbf{l}))\Pi_{CD}(\mathbf{l}) \\ &= a_S(\mathbf{l})\delta_S(\mathbf{l})\Pi_{CD}(\mathbf{l}) + a_{S'}(\mathbf{l})\delta_{S'}(\mathbf{l})r_{AB}(\mathbf{l})\Pi_{AB}(\mathbf{l})\Pi_{CD}(\mathbf{l}). \end{aligned} \tag{4.3}$$

It is interesting to notice how the presence of reflected rays contribute to enriching the plenacoustic view of the scene. For example, if the room includes scattering walls, the reflections are scattered in all directions, i.e. they are not originated from a single image source. Therefore the plenacoustic camera "sees" the diffuse reflections, originated from the reflector, coming from all directions. As a consequence the whole reflector's visibility region, representing all the rays originated from the reflector, will "brighten up". In this case the acquired ideal pleanacoustic image should be

$$\mathcal{I}^*(\mathbf{l}) = a_S(\mathbf{l})\delta_S(\mathbf{l})\Pi_{CD}(\mathbf{l}) + r_{AB}(\mathbf{l})\Pi_{AB}(\mathbf{l})\Pi_{CD}(\mathbf{l}).$$

The case of Lambertian (equip-diffusive) surfaces, however, is far more common in the optical case than it is in the acoustic one. However, we can always acoustically "illuminate" the environment from different locations in space to extract more information from the acoustic scene.

4.4. Array-based approximation of the plenacoustic camera

We now discuss how to acquire a plenacoustic image using a microphone array. In principle, just like in the optical domain, the plenacoustic camera can be thought of as an array of acoustic cameras, placed on a grid that samples the window of observation. Different setups are possible for this measurement procedure. If the acoustic scene is static and the signal emitted by the sources is stationary, the soundfield could be measured by simply moving an acoustic camera along the observation

window. If the acoustic scene is not static, then we need to resort to a one-shot acquisition procedure based on a spatially extended microphone array.

Implementing a plenacoustic camera with a microphone array means inferring the acoustic rays that intersect the array through space-time processing. Here we show how to build a soundfield camera that captures the plenacoustic function along a given segment, by using a microphone array that samples that segment. In order to do so, for each microphone position we will define a "local" acoustic camera. We then remap the collection of spatially distributed acoustic images onto the ray space to form the plenacoustic image. In the following paragraph we discuss in detail each step leading to the measurement of the plenacoustic image.

4.4.1. Single spatially extended array

The purpose of this paragraph is to illustrate a methodology for the acquisition of a plenacoustic image by means of a single microphone array. With reference to Figure 4.2 (a), we have a linear microphone array, whose sensors are at locations $\boldsymbol{x}_i$, $i = 1, ..., M$. Summarizing the overall procedure, we perform beamforming on compact sub-arrays within the microphone array and we measure the pseudospectrum (defined as the output power of the beamformer for different look directions) on each sub-array, which provides information about the distribution and intensity of acoustic rays passing through the acoustic center of the considered sub-array, i.e. its reference microphone. By performing beamforming on different sub-arrays we measure distribution and intensity of acoustic rays passing through a plurality of acoustic centers. Finally, the pseudospectra are visualized in the ray space, to obtain the plenacoustic image. A schematic overview of the plenacoustic image acquisition process is shown in Figure 4.2 (b).

Let us consider a sub-array centered at $\boldsymbol{x}_i$, i.e. the sensor located at $\boldsymbol{x}_i$ is the reference sensor of the considered sub-array. The sensors in the sub-array are located at $\boldsymbol{x}_j, j = i - \frac{W-1}{2}, \ldots, i + \frac{W-1}{2}$, where W is the number of microphones in the sub-array. The signals acquired by the sensors are $x_j(t), j = i - \frac{W-1}{2}, \ldots, i + \frac{W-1}{2}$. We apply a beamformer on these signals to obtain the spatial pseudospectrum

$$P_i(\theta) = \mathbf{h}^H(\theta)\hat{\mathbf{R}}_i\mathbf{h}(\theta), \tag{4.4}$$

where θ is the look direction, $\mathbf{h}(\theta)$ denotes the array weight vector, and $\hat{\mathbf{R}}_i$ denotes

the estimate of the autocorrelation matrix of the microphone signals

$$\mathbf{R}_i = E[\mathbf{X}_i(t)\mathbf{X}_i(t)^H],$$

where

$$\mathbf{X}_i(t) = [x_{i-\frac{W-1}{2}}(t), ..., x_{i+\frac{W-1}{2}}(t)]^T,$$

and

$$\hat{\mathbf{R}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_i(t)\mathbf{X}_i(t)^H.$$

In general, the pseudospectrum $P_i(\theta)$ can be measured using any beamforming technique. However, discussion about beamforming methods is outside the scope of this thesis. For more details see the Appendix C. The derivation of the pseudospectrum is repeated for all the possible sub-arrays, i.e. for $i = (W+1)/2, \ldots, M - (W+1)/2$.

In order to map $P_i(\theta)$ onto the ray space, notice that an acoustic ray with travel direction θ and captured by the sub-array whose reference microphone is at $\boldsymbol{x}_i = [x_i, y_i]^T$ has line parameters

$$\begin{gathered}\mathbf{l}(\theta, x_i, y_i) = k\left[-\cos(\beta), \sin(\beta), \cos(\beta)x_i - \sin(\beta)y_i\right]^T, \\ k > 0,\ \beta = \theta - \alpha + \pi/2\ ,\end{gathered} \tag{4.5}$$

where α is the orientation of the array, as shown in Figure 4.2 (a). Equation(4.5) is the conversion rule in order to go from $P_i(\theta)$ to $P_i(\mathbf{l})$. Notice that as θ varies $\mathbf{l}(\theta, x_i, y_i)$ describes a line in the ray space, i.e. the line $\overline{\mathbf{x}}_i$ corresponding to $\boldsymbol{x}_i$. Furthermore, due to the fact that $\boldsymbol{x}_i$ lies on the line of the array, the line $\overline{\mathbf{x}}_i$ belongs to the visibility region of the plenacoustic camera and passes through the intersection point between $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_M$ as shown in Figure 4.2 (c). Finally, the plenacoustic image acquired by the microphone array is the collection of the sub-array pseudospectra represented in the ray space, i.e.

$$\begin{gathered}\mathcal{I}(\mathbf{l}) = \{P_i(\mathbf{l}) | \mathbf{l} = \mathbf{l}(\theta, x_i, y_i), \\ i = \tfrac{W+1}{2}, \ldots, M - \tfrac{W+1}{2}, 0 < \theta < \pi\}\ .\end{gathered} \tag{4.6}$$

The use of the spatial windowing W allows us to trade between conflicting needs:

- ▷ the use of extended arrays enables a wider region in the ray space to be visible;

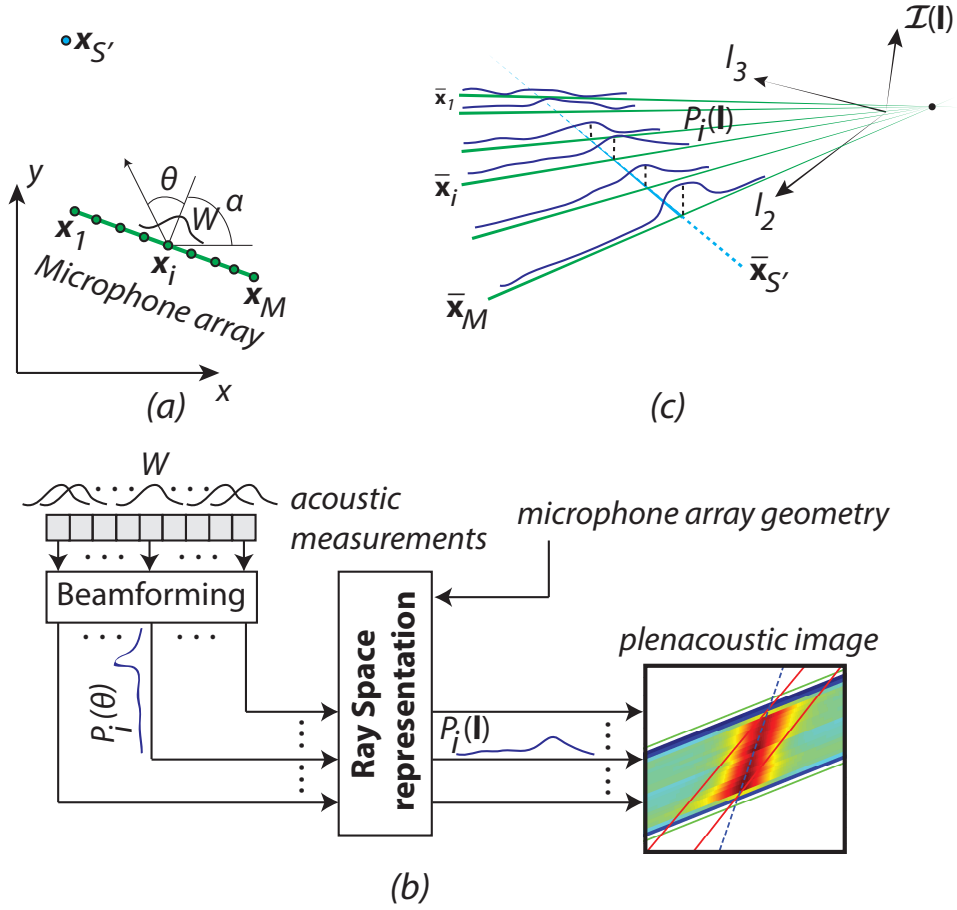


Figure 4.2.: (a) The microphone array; (b) Schematic overview of the plenacoustic image acquisition process; (c) The ray space representation of the plenacoustic measurements – the plenacoustic image.

▷ in order to estimate rays that pass through a point $\boldsymbol{x}_i$ of the array, the array must be sufficiently compact in space.

In other words, the complete array has to be "big" with respect to the scene of interest (near-field) in order to get a "panoramic view" of the acoustic scene; whereas the sub-arrays have to be "small" with respect to the scene of interest (far-field) in order to approximate point-like directional receivers. Spatial windowing overcomes this problem but introduces an *aperture* phenomenon and other distortions. For example, a smaller spatial window reduces the angular resolution (it "blurs" the plenacoustic image) but increases the spatial resolution (information about rays passing through a single point $\boldsymbol{x}_i$). On the other hand, a larger spatial window improves the angular

resolution (assuming the far-field assumption made by most beamforming techniques is still valid) but decreases the spatial resolution. On the positive side, windowing suggests that an array of many microphones could be replaced with a smaller array that slides along its axis (under the hypothesis of stationary soundfield).

The resulting plenacoustic image $\mathcal{I}(\mathbf{l})$ is a sampled and blurred version of the ideal pleanacoustic image $\mathcal{I}^*(\mathbf{l})$ (4.1) as illustrated in Figure 4.2 (c). More specifically, the fact that the image is sampled means that $\mathcal{I}(\mathbf{l})$ samples $\mathcal{I}^*(\mathbf{l})$ over the rays $\mathbf{l}$ that are bound to pass through microphone positions $\boldsymbol{x}_i, i = (W+1)/2, \ldots, M-(W+1)/2$, i.e. over the lines $\overline{\mathbf{x}}_i$ in the ray space. Conversely, the blurring comes from the fact that the image $\mathcal{I}(\mathbf{l})$ represents the ideal image $\mathcal{I}^*(\mathbf{l})$ convolved with the aperture function of the given beamforming technique. We can thus write

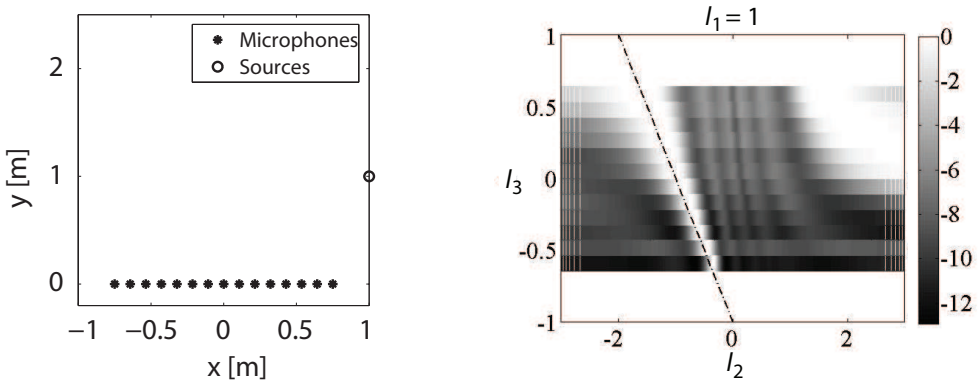
$$\mathcal{I}(\mathbf{l}) = \left\{ \int_{\mathbf{l}' \in \overline{R}_{CD}} \mathcal{I}^*(\mathbf{l}') p_{\mathbf{l}'}(\mathbf{l}-\mathbf{l}') d\mathbf{l}' \mid \mathbf{l} = \mathbf{l}(\theta, x_i, y_i), \right. \\ \left. i = \tfrac{W+1}{2}, \ldots, M - \tfrac{W+1}{2}, 0 < \theta < \pi \right\}, \tag{4.7}$$

where $p_{\mathbf{l}'}(\mathbf{l}-\mathbf{l}')$ is the aperture function and the subscript $\mathbf{l}'$ indicates that, in general, it is not constant for all directions $\mathbf{l}'$.

The image configuration depends on the geometry (reflector and source positions) while the amplitudes vary in time and frequency according to the source signal, radiation pattern of the source, the polar pattern of the microphones, the reflection coefficients and traveling path distance. In case of wideband sources we can obtain a number of images for different frequency bands of interest or (being interested in extracting only geometric information) a single image combining images at different frequencies. In particular, in Chapter 5 we use the wideband Minimum Variance Distortionless Response (MVDR) beamformer [75] for the localization of acoustic sources and reflectors. On the other hand, in Chapter 6, where we are interested in extracting the frequency dependent reflection coefficient, we apply the delay-and-sum (DAS) beamformer [76], [77]. A detailed discussion about the beamforming techniques is outside the scope of this dissertation. A short description of both DAS and MVDR techniques can be found in the Appendix C. However, various alternatives could be employed. In order to improve accuracy and resolution, other robust and superdirective methods can be used, as suggested in [78].

Although the acquisition process could be easily performed with an arbitrarily shaped array, the linear array has the advantage of maximal extension and, consequently, best field of view. It also has the advantage of being easily manageable

in the ray space. We should bear in mind, however, that the field of view of this camera is limited to half of the geometric space (in front of it).



(a) (b)

Figure 4.3.: (a) An acoustic scene with a ULA plenacoustic camera and a source $\boldsymbol{x}_S = [1, 1]^T$ in the geometric space; (b) The corresponding plenacoustic image $\mathcal{I}(\mathbf{l})$[dB], whose amplitude has been normalized and expressed on a dB scale, shown in the reduced ray space with $l_1 = 1$.

Figure 4.3(a) shows an acoustic scene that includes a Uniform Linear Array (ULA) of 15 microphones spaced of 0.11 m. An acoustic source placed in $\boldsymbol{x}_S = [1, 1]^T$ produces a pass-band signal whose spectrum lies between 300 Hz and 10 kHz. The corresponding simulated plenacoustic image obtained using wideband MVDR is shown in Figure 4.3(b). For clarity of visualization, the resulting image is displayed after order-zero interpolation (piecewise constant), in order for $\mathcal{I}(\mathbf{l})$ to be continuous with respect to microphone sampling. The line of Figure 4.3(b) is the dual $\overline{\mathbf{x}}_S$ of the source, i.e. the representation of the source in the ray space. As we can see, the plenacoustic image $\mathcal{I}(\mathbf{l})$ exhibits a ridge in the same location as $\overline{\mathbf{x}}_S$. This ridge is, in fact, a blurred version of the visible portion of the dual of the source and the magnitude of the blurring varies with both l_2 and l_3. This is due to the fact that a ULA does not exhibit a uniform resolution over θ [79]. As we can see in Figure 4.3(b), in particular, the farther the point from the source, the larger the incidence angle, the greater the blurring. This blurring in the plenacoustic image causes a loss of resolution, which prevents us from being able to tell multiple acoustic objects apart when they lie too close to each other.

In the lower left area of the plenacoustic image in Figure 4.3(b) we also notice a rather large bright area, caused by aliasing. The signal emitted by the source, in fact, has frequency content that goes beyond the spatial Nyquist frequency. This

phenomenon will be better characterized later on in the next section. We will also show that plenacoustic imaging is, in fact, robust with respect to aliasing.

4.4.1.1. Angular Aliasing

Aliasing is a well known phenomenon in space-time processing, which causes an error in the localization of the acoustic source. An aliased pseudospectrum exhibits multiple lobes of comparable magnitude, known as grating lobes [79], which are replicas of the main lobe. In order to prevent aliasing, the distance d between adjacent sensors needs be kept sufficiently small.

As far as ULAs are concerned, the no-alias condition is $d < \lambda/2$, where λ is the wavelength corresponding to the maximum frequency contained in the signal and d is the distance between adjacent sensors. We focus here on the impact of spatial aliasing on plenacoustic images. The presence of aliasing depends only on the deployment of sensors. As an exact characterization of aliasing for MVDR is not possible in a theoretical fashion, we present analytical results for the case of delay-and-sum beamformer [79], which can be generalized with some approximation to the MVDR.

Let us consider the ith subarray, whose central microphone is in $\boldsymbol{x}_i = [x_i, y_i]^T$. The angle under which this sub-array sees the acoustic source in $\boldsymbol{x}_S = [x_S, y_S]^T$ is

$$\theta_i = \arctan\left(\frac{y_i - y_S}{x_i - x_S}\right) .$$

The acoustic source produces a "monochromatic" signal (a single tone) of wavelength λ. For the delay-and-sum beamformer, the contribution of the sub-array to the plenacoustic image is

$$P_i(\theta) = C\frac{\sin\left[\frac{\pi W d}{\lambda}(\sin\theta - \sin\theta_i)\right]^2}{\sin\left[\frac{\pi d}{\lambda}(\sin\theta - \sin\theta_i)\right]^2} , \tag{4.8}$$

where C is a positive constant. Spatial aliasing occurs when the denominator is zero, i.e. when

$$\frac{\pi d}{\lambda}(\sin\theta - \sin\theta_i) = m\pi,\ m \in \mathbb{Z} ,$$

which gives

$$\theta = \arcsin\left(\frac{n\lambda}{d} + \sin\theta_i\right) , -\pi/2 \leq \theta < pi/2 , n \in \mathbb{Z} \tag{4.9}$$

Although eq. (4.9) is derived for the case of the delay-and-sum beamformer, the same equation can be used, with some degree of approximation, also for other beamforming techniques.

Figure (4.4) shows the same plenacoustic image of Figure 4.3(b), where a continuous line marks the location of the grating lobes as predicted with equation (4.9), with $n = -1$. As we can see, although derived for the delay-and-sum beamformer, the curve well approximates the location of the grating lobes also in the case of MVDR beamforming. Small crosses denote the location of aliasing peaks, as detected with a peak-picking algorithm. Notice that such peaks are not collinear. This means that any line detection tool such as the Hough transform would allow us to easily discriminate between peaks related to real sources and peaks due to grating lobes.

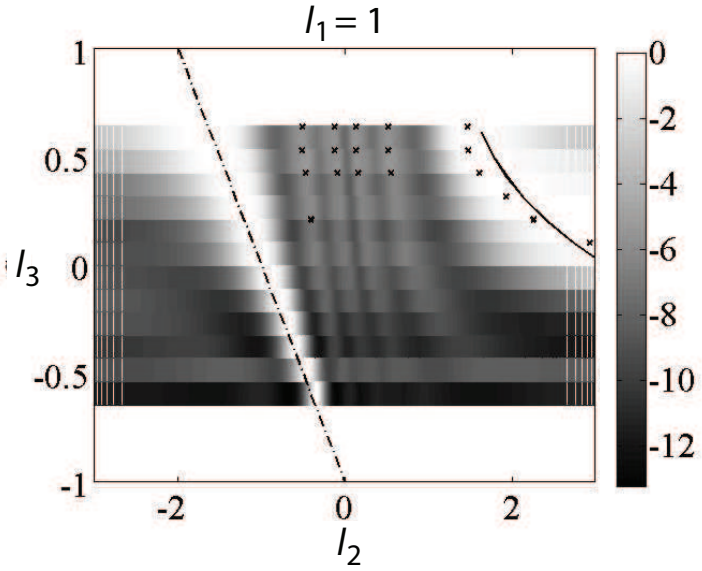


Figure 4.4.: Example of plenacoustic image with aliasing. The geometric setup is the same as in Fig. 4.3. The dashed line is the dual of the source; crosses mark the detected aliasing peaks; the continuous line, computed from (4.9) approximately predicts the location of aliasing peaks.

4.4.1.2. Resolution

The resolution is defined as the minimum angular distance between two sources that makes the related pseudospectrum peaks discernible. For notation simplicity we consider the configuration in Figure 4.3 in which the l_3 axis corresponds to the spatial variation (position of the reference microphone $\boldsymbol{x}_i$) whereas the l_2 axis

corresponds to the angular variation (direction of arrival θ). As far as the spatial resolution is concerned, the resolution is limited by the number of pseudospectra $P_i(\theta)$ contributing to the plenacoustic image. An increase of the resolution on the l_3 axis can be obtained decreasing the number W of microphones in each sub-array. The cost is, however, the decrease of the resolution on the l_2 axis.

We now consider the angular resolution. The discriminating ability depends on the adopted peak-picking algorithm, therefore it is more of an operative definition. In this work the resolution is evaluated by sizing the width of the lobe of the pseudospectrum corresponding to the Direction Of Arrival (DOA) of a point-like source, i.e. we characterize it as a point-spread function. Let $l_2^{\max}$ be the value of l_2 corresponding to the DOA of the source and $\delta l_2^+ > 0$ be the interval on the l_2 axis such that

$$\mathcal{I}(1, l_2^{\max} + \delta l_2^+, l_3)|_{\mathrm{dB}} = \mathcal{I}(1, l_2^{\max}, l_3)|_{\mathrm{dB}} - \Delta \, ,$$

Δ being a given threshold. Similarly we define $\delta l_2^- < 0$ as the value of l_2 such that

$$\mathcal{I}(1, l_2^{\max} + \delta l_2^-, l_3)|_{\mathrm{dB}} = \mathcal{I}(1, l_2^{\max}, l_3)|_{\mathrm{dB}} - \Delta \, .$$

Finally, we define the width of the lobe as

$$\delta l_2 = \delta l_2^+ - \delta l_2^- \, , \tag{4.10}$$

which clearly depends on Δ. For the applications presented in this dissertation we preliminarily verified that the peak picking algorithm adopted in this manuscript requires $\Delta \geq 4$ dB in order to discriminate between peaks in the pseudospectra associated to multiple sources.

A closed-form expression of δl_2 can only be found for the delay-and-sum beamformer, and not for MVDR beamformer. This is why we performed simulations. Figure 4.5 plots δl_2 for $W = 3$ and $W = 5$ for various source positions. More specifically, the source is placed at a distance of 1.5 m, and the angle θ formed by the source and the y axis varies between 0° and 45°. The microphone array has the same configuration of Figure 4.3. For visualization convenience, the value of δl_2 has been converted in angles, as the range of variability of l_2 is too large for a clear representation. Notice that for $\Delta = 8$ dB there are angles for which $\delta l_2|_{\mathrm{deg}} = \pi$. In this situation δl_2^+ is not defined, as the lobe of $\mathcal{I}(1, l_2, l_3)|_{\mathrm{dB}}$ does not decrease to $\mathcal{I}(1, l_2^{\max}, l_3)|_{\mathrm{dB}} - \Delta$ for $m > m_{\max}$. Notice that there is no significant improvement

on the resolution for $W = 3$ and $W = 5$ when $\Delta = 4$ dB.

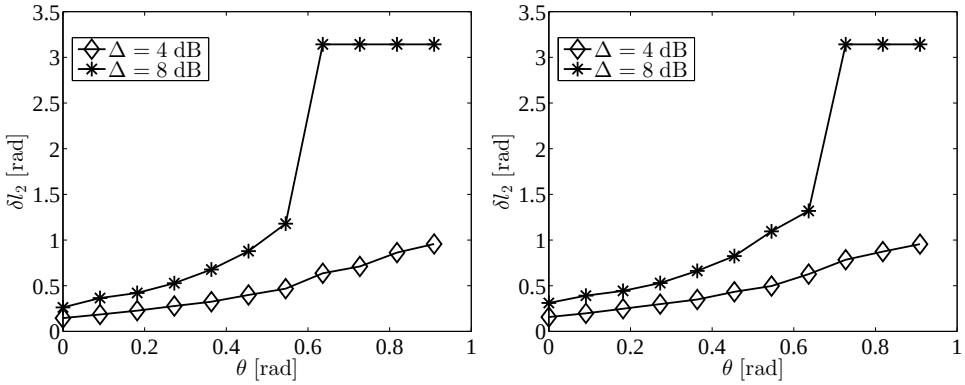


Figure 4.5.: Resolution on the l_2 axis as the angle of the source varies: (a) $W = 3$; (b) $W = 5$.

4.4.1.3. Spatial Sampling

Real plenacoustic images differ from the ideal ones not only for their limited resolution, but also because they are obtained by sampling the observation window on a finite number of points corresponding to the acoustic centers of the sub-arrays. As suggested above, increasing W, and therefore the resolution on the l_2 axis has the negative consequence of decreasing the number of samples that form the plenacoustic image on the l_3 axis. The optimal tradeoff between resolution and spatial sampling issues depends on the specific application.

4.4.2. Spatial distribution of small compact arrays

An acoustic camera (in a more traditional sense) could be seen as a compact array of microphones (whose size is negligible compared with the size of the imaged scene). This camera would be able to estimate the magnitudes of all rays that pass through a single point in space $\boldsymbol{x}_i$ (the location of the array), therefore its field of view in the ray space corresponds to a single line $\overline{\mathbf{x}}_i$. Along this line we find the values of the pseudospectrum $P_i(\mathbf{l})$ corresponding to the directions of arrival of the rays. Examples of acoustic cameras are the cylindrical arrays (2D case) and spherical arrays (3D case). An alternative and more expensive implementation of a plenacoustic camera could be a spatial distribution of small acoustic cameras of

this sort (e.g. Eigenmikes), each taking care of estimating sound field magnitudes along all directions. This would have the effect of reducing the aperture problem; increasing the field of view (to a full angle); distributing the processing (over the small acoustic cameras) and compacting the information (each small acoustic camera would transmit a set of vectors and the related amplitudes). The expensive hardware requirements of this scenario could be overcome in the near future thanks to the availability of inexpensive integrated microphone arrays. This approach resembles the construction of some plenoptic cameras, which can be built as arrays of mini-cameras. With this perspective in mind, the use of the proposed method combined with robust superdirective beamforming techniques (such as [78]) becomes particularly interesting.

4.5. Perspective work: extension to 3D

In this Chapter we considered acoustic images for 2D geometries. Notice however that the approach can be extended to 3D environments using a planar array and the 3D parametrization described in Chapter 2. We have seen in Chapter 3 that the generalization of the modeling procedure to the 3D case is not a trivial task. Given the 3D parametrization described in Chapter 2, however, the generalization of the analysis steps should be simpler. In order to construct theoretically the visibility regions we have to take into account the intersection with the non-convex Plucker quadric in order to guarantee that points in $\mathbb{P}^5$ correspond to real rays in $\mathbb{R}^3$. On the other hand, if we map the real acoustic measurements into the ray space, they will lie automatically on the surface of the Plucker quadric and, therefore, no intersection is needed. Furthermore, although the data lies on the quadric surface, the boundaries between different primitives are still represented by linear constraints in $\mathbb{P}^5$ (see (2.11) and (2.14)). The planar array is subdivided in sub-arrays. For each sub-array the angular distribution of acoustic power is estimated through the computation of a pseudospectrum and the collection of pseudospectra represented in the ray space forms the plenacoustic image similarly as done in (4.6). To conclude, we can say that it is possible to generalize the proposed approach to 3D geometries and the considerations we made have general validity.

4.6. Conclusive remarks

In this Chapter we proposed a novel approach to acoustic scene analysis based on the concept of plenacoustic imaging. We first defined the plenacoustic camera as a device that captures the acoustic radiance along all the acoustic rays that cross an observation window. After discussing the process of ideal plenacoustic image formation we introduced a possible implementation based on microphone array, and discussed its behavior in terms of resolution and aliasing. The larger the number of microphones of the array, the greater the detail in the acquired images. Recent progress in MEMS and integrated electronics technology suggests that the number of microphones that can be managed in integrated arrays is on a growing trend. Plenacoustic imaging can therefore become particularly useful for managing and organizing the massive data that such devices will be able to collect.

One key aspect of this approach is that we are defining a single layer of space-time processing (required for generating the plenacoustic images), which can be shared "as is" by a wide variety of applications. With the ray space parametrization described in the Chapter 2, acoustic primitives such as sources and reflectors, are mapped onto linear features/regions of the plenacoustic image, which greatly simplifies acoustic scene analysis algorithms. In fact, the plenacoustic imaging is a powerful analysis paradigm for multiple reasons:

- it turns problems of acoustic analysis in space and time into problems of pattern analysis on images, which can be approached with methods found in the rich literature of pattern analysis, computer vision and multidimensional signal processing;
- image generation becomes a pre-processing step that remains the same throughout a wide range of applications and is highly parallelizable, thus paving the way to the production of a shared hardware framework;
- objects in the acoustic scene correspond to the image patterns that are easily discerned and modeled, which simplifies pattern analysis/detection/extraction, and enables super-resolution methods. Our definition of the ray space, in particular, makes such patterns linear, with clear advantages in terms of detection performance.

The applications addressed in the following Chapters have the purpose of offering an initial proof of concept of these points and will be further explored and expanded

in future works. In particular the Chapter 5 will focus on two specific applications: multiple near-field source localization and reflector localization. These problems have been addressed numerous times in the literature. For example a near-field bemaforming method for the localization of acoustic sources is proposed in [80]. Reflector localization methods were proposed in [81] and [5]. These, however, were ad-hoc solutions devised for the specific problem at hand. Plenacoustic imaging, on the other hand, gives us the possibility to adopt a common framework that can be reused throughout an even wider range of problems.

We believe that this approach to the analysis could enable the development of novel solutions for a wide class of applications such as wave field analysis/extrapolation; image fusion; image-based self-calibration; source separation; environment inference, etc.

Chapter 5

Application to environment awareness: geometry inference

In this Chapter we use the plenacoustic images for geometry inference. In particular we "look" at the acoustic scene, i.e. obtain an overview of the sound field distribution in different positions in space, by taking a number of plenacoustic images. The resulting plenacoustic images are then analyzed in order to reconstruct the geometry of the environment. Therefore we are imitating, to a certain extent, the procedures used in computer vision where a number of images, taken from different positions in space, are used for the geometry reconstruction. In particular, knowing how the acoustic counterparts of geometric objects should occur in the ray space, we can search the pleanacoustic images for the corresponding patterns. Whereas the presence of multiple acoustic sources can be difficult to deal with in traditional systems, using the plenacoustic images the contributions of individual sources are easily separated and all the sources (up to the resolution limit determined by the non ideal acquisition process) are localized simply by finding lines on the image. Once the positions of direct and the image sources are estimated, the positions and the orientations of acoustic reflectors are determined by finding "mirrors" that, given the direct source position, produce the corresponding images sources. Furthermore, while the traditional systems, based on

small compact microphone arrays, estimate just the surface on which the reflector lies, the spatially distributed observations capture also the visibility discontinuities that can be used to estimate the reflector extension. In particular, in the ray space the image source lines are "windowed" by the visibility region of the reflector that generates the reflection. If the edges of this visibility region are acoustically visible by the plenacoustic camera the extension of the corresponding reflector can be estimated. Summarizing: line detection corresponds to source localization while the edge detection corresponds to the reflector endpoint estimation. We finally show a set of simulations and acoustic experiments that validate the feasibility and effectiveness of the proposed method.

5.1. Motivations

Motivated by the recent progress in environment-aware audio processing, the interest in the estimation of the acoustic properties of the environment is steadily growing [1]. Information on the environment, in fact, can be used for boosting the performance of space-time audio processing algorithms promising to significantly push the boundaries of audio signal processing. For example, it can be successfully exploited in the localization and characterization of sources in reverberant environments [82]; or for the compensation of undesired reflections in wavefield rendering applications [83]. However, environment-aware sound processing relies on knowledge of the geometry and the acoustic proprieties of the environment, which need to be measured or estimated in advance. Although the reconstruction of the environment geometry from a series of images is a well known problem in computer vision, what is visible in optics is not necessary visible acoustically and vice versa (think about small objects, transparent surfaces or low reverberant walls). This motivates the acoustic probing and sensing in order to estimate the position of reflectors. However, the computer vision procedure in not easily extended to acoustics.

In order to address this need, a number of techniques for the localization of reflective surfaces based on acoustic measurements have begun to appear in the literature. These solutions, usually use the acoustic measurements (possibly along with prior information) to define nonlinear constraints on the location of the reflectors. Multiple constraints (each associated to a different position of the source or the microphone)

are combined in a single cost function, whose minimization yields the estimated line (2D case) or plane (3D case) where the reflector lies. The methods proposed in the literature are generally very accurate but tend to differ from each other in assumptions made and hardware used making some of them more suitable for certain scenarios than others.

One problem that often arises with the above iterative minimization process is associated to the inherent nonlinearity of the cost function, which causes the optimization to get easily trapped into local minima. In addition, when multiple reflectors are in place, a preliminary clustering process is needed in order to understand which reflector has generated which acoustic event, which tends to make the method more complex to implement and more prone to errors. In this Chapter we propose a method that aims at overcoming these limitations and extracting richer information on the environment. In particular, we want to turn the reflector estimation process into a linear one. In order to do so, acoustic measurements have to be represented in a space in which the reflector is described by linear constraints. We also want to be able to manage multiple reflectors at once, therefore we need acoustic measurements to be organized and represented in a space that exhibits a high degree of regularity. This means that acoustic events coming from different reflectors need to be easily discernible. With the method that we propose we also address two other important needs: robustness against incomplete measurements; and level of detail in the extracted information. Incomplete measurements are frequently due to the fact that a reflector might turn out to be “visible" from only some of the microphones. As for the richness of the extracted information, we would like to be able to estimate not just the location of each reflector but also its spatial extension.

In order to address all of the above requirements, we propose a space-time processing algorithm conceived for microphone arrays whose spatial extension is comparable with the operating distance (near-field operation). This algorithm is based on the concept of the plenacoustic image introduced in the Chapter 4. In fact, the ray space parametrization, used for the representation of the plenacoustic measurements (i.e. the spatially distributed measurements of the soundfield), turns out to be quite effective as the geometric primitives of interest (acoustic sources, microphones, reflectors) are all represented in this ray space by regions that can be clearly identified and easily described by one or more linear constraints. As a consequence, localizing sources and reflectors becomes a linear procedure. Furthermore, the increase of dimensionality in the representation of the measurements from acoustic images to

plenacoustic images paves the way to the estimation of the extension of the reflector, as it allows us to observe the transition point between the absence and the presence of the reflective path coming from the reflector under study, which is necessary, in turn, for the estimation of the extension of the reflector. In fact, most of the techniques in the literature for geometry environment estimation use compact arrays, which present, however, the disadvantage of sampling the soundfield in a limited portion of space, thus enabling the estimation of the only portion of the environment that is visible from the array location. The compact arrays either observe the acoustic event generated by the reflector or do not observe it at all while, in order to estimate the reflector extension, the measurement system should be able to observe the discontinuity in the presence of an acoustic event generated by the reflector.

5.2. Related work

The problem of finding reflective surfaces in the environment has recently been addressed by a number of authors. The proposed methods were developed mostly for 2D geometries with possible extensions to 3D. All the methods assume valid the hypothesis of optical acoustics, i.e. reflections undergo the Snell's law. Usually only the first most significant reflections are considered. However, they differ in equipment used and assumptions.

Most techniques estimate the reflector positions matching the estimated Room Impulse Responses (RIRs) with template reflections or building constraints from estimated Times Of Arrival (TOAs) or Directions Of Arrival (DOAs). However, all the methods suppose walls to be infinite and as a consequence the reflections to be always visible by the acoustic system.

In [36] a loudspeaker moves on a circular trajectory within a 2D space and emits the controlled noise acquired by a single microphone. A likelihood map is built as a correlation between acquired signal and a template signal obtained by simulating the propagation of the signal with respect to all the potential obstacle locations. The reflector position is found in correspondence of the maximum of the likelihood map. In [37] the necessity for a priori knowledge of the source signal is removed developing a method based on inverse mapping of the multi path propagation model. The reflective surface is localized as the maximum of the pseudo likelihood map defined as a spatial combination of generalized crosscorrelation functions between signals received by a microphone pair.

Authors in [35] propose a method based on a minimization of a cost function obtained as a combination of a number of quadratic constraints. A single microphone and a synchronized moving loudspeaker are used in a 2D space to obtain constraints based on TOAs. A knowledge of a probing signal allows extraction of a RIR from the acquired signal applying a crosscorrelation operation. Maxima of the impulse response give TOA measurements that define the locus of candidate reflection points as an ellipse. Different source positions define different elliptic constraints and the reflector is found as a line of tangency to all the ellipses. In case of multiple reflectors the TOAs are labeled by means of a generalized Hough transform. This approach is modified in [4] to account for uncontrolled emissions and unknown source location. A blind system identification technique is applied in order to extract time differences of arrival (TDOAs) associated with direct paths. Successively TDOAs are used to estimate the source-microphone distance and as a consequence the TOAs. In case of multiple reflectors the related ellipses are grouped iteratively. The approach is further extended in [84] with an analytic (closed form) estimator and a second stage correction based on a Hough transform for a more robust solution. Another improvement is provided in [85], where the optimization problem is turned into a linear least-squares problem. Finally, the localization of planar reflectors is approached in [40], [41], [86]. All the above techniques use TOAs as acoustic measurements. In [5] a different type of quadratic constraint is obtained using DOAs related to reflective paths. DOA estimates are obtained using a circular microphone array and applying a wideband Minimum Variance Distortionless Response (MVDR) algorithm [75]. With DOAs the reflector is bound to be tangential to a parabola having the focus in the source location and the measured DOA as directrix. Once again, multiple constraints are obtained for multiple positions of the source. Notice that, in order to set up the parabolic constraint, the source location needs to be known in advance.

In order to make reflector localization more robust to interfering sources present in the room and avoid measurement of RIRs (in case of the unknown source signal) the authors in [78] propose the use of a robust MVDR beamformer [87], focusing matrices and frequency smoothing [75], [88] for DOA estimation and the successive extraction of signals originating from estimated DOAs by means of a superdirective RLS-FIB beamformer [89]. Using the crosscorrelation analysis on extracted signals the TDOA estimates are obtained. Finally, in combination with the information about the position of the direct source, the reflector location is estimated. A sim-

ilar procedure is accomplished in three-dimensional environments in [81], using a spherical microphone array and the EB-MVDR beamformer [90].

It is also worth mentioning the work in [38] where authors analyze the (invertible) mapping between geometry of a convex polygonal 2D room and a single RIR and develop an algorithm that retrieves the room geometry from the measured RIR. Finally in [39] authors propose the method for reflector localization in simple 3D environments by means of a compact microphone array and an integrated and synchronized loudspeaker that probes the environment with a known test signal. The estimated RIRs are fitted by a non-parametric method with synthetically generated reflections. Reflections are validated and classified as 1st, 2nd and 3rd order reflections and used for reflector localization.

5.3. Problem formulation

Consider the scenario in Figure 5.1, where a linear microphone array is installed in an environment. The acoustic scene is composed by the source $\boldsymbol{x}_S$ and the reflector $\boldsymbol{x}_A\boldsymbol{x}_B$. We consider reflectors in the environment to be linear. Along with the direct signal coming from the source $\boldsymbol{x}_S$, sensors in the array capture also a delayed and attenuated replica of it, coming from the specular reflection from $\boldsymbol{x}_A\boldsymbol{x}_B$. According to the Snell's law reflections can be thought as originated from image sources obtained by mirroring the source with respect to reflectors (image source $\boldsymbol{x}_{S'}$ shown in Figure 5.1 (a)). Our objectives are: estimate the location of the reflective surface; and estimate the location of its endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$, under the assumption that they are acoustically "visible" from the microphone array. Due to the specular reflection law, in fact, only a portion of space is relevant to the reflective path. More specifically, the visibility of the image source $\boldsymbol{x}_{S'}$ is limited in space by the extension of the reflective surface. In fact the rays originated from the image source are bound to intersect the reflector, as illustrated in Figure 5.1(a). In order to estimate the endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ of the reflector we observe the discontinuity of the presence of the reflective path when $\boldsymbol{x}_S$ is moved at different positions in space.

A schematic overview of the algorithm is shown in Figure 5.1 (b). The signals coming from the array are processed and the plenacoustic image in the ray space is obtained. From the plenacoustic image we are then able to estimate direct and image source locations, which are then used to estimate the reflector location. When the environment is probed from multiple source locations we are also able to estimate

the reflector endpoints. To this purpose the algorithm uses both plenacoustic images and source locations.

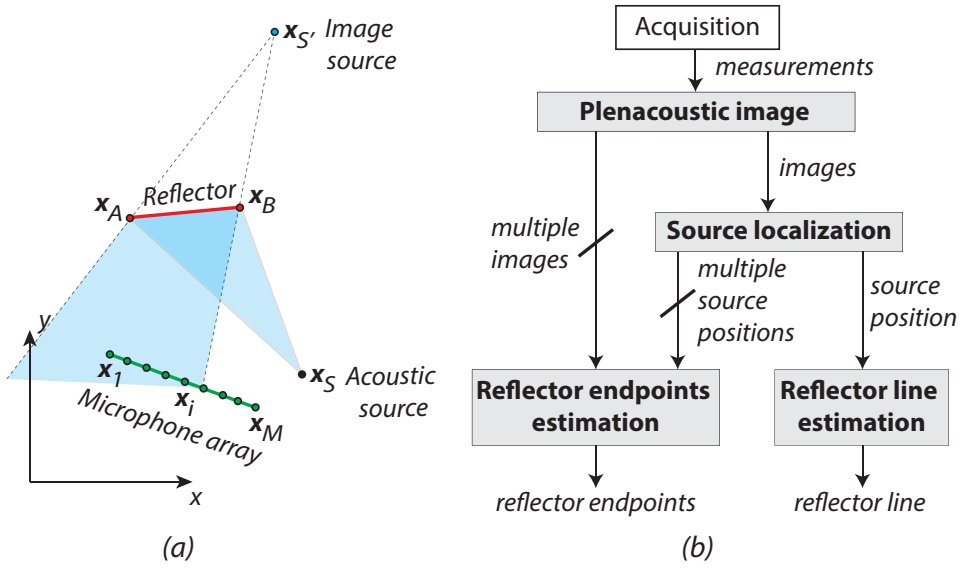


Figure 5.1.: (a) Problem setup: a microphone array captures the soundfield produced by an acoustic source within an environment with a single reflective surface; (b) Steps of the reflector estimation procedure.

In the next sections we describe the algorithms for geometry reconstruction using plenacoustic images. We start from the case of a single reflective surface. As described in the block diagram in Figure 5.1, in order to localize the reflector we need beforehand to localize the acoustic source. At this regard, we initially devise an algorithm for the localization of a single source through plenacoustic images, either the direct or the reflected one. In order to localize the reflector, we then extend the methodology to the case of multiple sources. Finally, we propose the technique for the localization of multiple reflectors along with their extensions.

5.4. Source localization

In this paragraph we discuss the localization of acoustic sources from plenacoustic images $\mathcal{I}(\mathbf{l})$ (4.6). We will proceed with the example of localizing the image source $\boldsymbol{x}_{S'}$ in Figure 5.2 (a). Notice, however, that the same algorithm could be used for localizing the direct source. In Figure 5.2 (a) only $N_{S'}$ microphones sense the reflective path from $\boldsymbol{x}_{S'}$. The set $\mathcal{V}_{S'}$ of microphones $\boldsymbol{x}_i$ that sense the reflective paths originated from the image source $\boldsymbol{x}_{S'}$ is found by considering only the pseudospectra

that exceed a predetermined threshold β, i.e.

$$\mathcal{V}_{S'} = \{i \mid \max(P_i(\mathbf{l})) > \beta \max(\mathcal{I}(\mathbf{l}))\}. \tag{5.1}$$

We then localize the maxima of the pseudospectra $P_i(\mathbf{l}), i \in \mathcal{V}_{S'}$, i.e.

$$\hat{\mathbf{l}}_{S'}^{(i)} = \arg\max_{\mathbf{l}}(P_i(\mathbf{l})),\ i \in \mathcal{V}_{S'}. \tag{5.2}$$

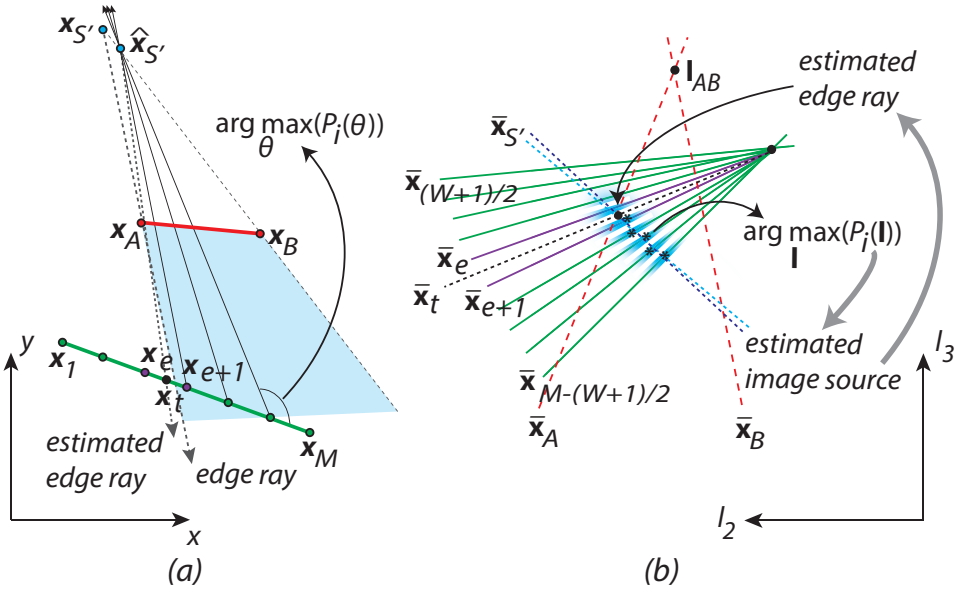


Figure 5.2.: (a) Configuration in the geometric space; (b) Configuration in the reduced ray space: linear regression is used to estimate the image source position $\boldsymbol{x}_{S'}$; the edge ray $\mathbf{l}_A$ is found at the intersection between the image source line $\overline{\mathbf{x}}_{S'}$ and the edge of the reflector's visibility region $\overline{\mathbf{x}}_A$.

The rays $\hat{\mathbf{l}}_{S'}^{(i)}$, $i \in \mathcal{V}_{S'}$, are originated from the image source $\boldsymbol{x}_{S'}$ and, therefore, they should all stay on the line $\overline{\mathbf{x}}_{S'}$ in the reduced ray space, as depicted in Figure 5.2 (b). Thus, using (2.2), we can write

$$\hat{\mathbf{l}}_{S'}^{(i)T}\mathbf{x}_{S'} = 0,\ i \in \mathcal{V}_{S'}, \tag{5.3}$$

where $\mathbf{x}_{S'}$ are the homogeneous coordinates of $\boldsymbol{x}_{S'}$. We write the set of $N_{S'}$ equations (5.3) in a matrix form as

$$\hat{\mathbf{M}}_{S'}\mathbf{x}_{S'} = 0, \tag{5.4}$$

where the matrix $\hat{\mathbf{M}}_{S'} = [\hat{\mathbf{l}}_{S'}^{(i_1)}, \hat{\mathbf{l}}_{S'}^{(i_2)}, ..., \hat{\mathbf{l}}_{S'}^{(i_{N_{S'}})}]^T$, $i_1, i_2, ..., i_{N_{S'}} \in \mathcal{V}_{S'}$ has dimensions $N_{S'} \times 3$. If we write $\hat{\mathbf{M}}_{S'} = [\hat{\mathbf{H}}_{S'}, -\hat{\mathbf{d}}_{S'}]$ we can reformulate equation (5.4) as

$$\hat{\mathbf{H}}_{S'} \boldsymbol{x}_{S'} = \hat{\mathbf{d}}_{S'} \ , \tag{5.5}$$

where $\hat{\mathbf{H}}_{S'}$ and $\hat{\mathbf{d}}_{S'}$ have dimension $N_{S'} \times 2$ and $N_{S'} \times 1$, respectively. Finally, we estimate the Cartesian coordinates of the image source position $\boldsymbol{x}_{S'}$ as the least-squares (LS) solution of (5.5),

$$\hat{\boldsymbol{x}}_{S'} = (\hat{\mathbf{H}}_{S'}^T \hat{\mathbf{H}}_{S'})^{-1} \hat{\mathbf{H}}_{S'}^T \hat{\mathbf{d}}_{S'}. \tag{5.6}$$

The source localization turns out to be a linear problem in the ray space as it consists in fitting a strait line to a set of points representing rays originated from the acoustic source. This procedure corresponds in the geometric space to finding the intersection between all the lines going from microphones $\boldsymbol{x}_i$ with the corresponding DOAs $\hat{\theta}_i = \arg\max_{\theta}(P_i(\theta))$, $i \in \mathcal{V}_{S'}$, as illustrated in Figure 5.2 (a).

5.5. Reflector line estimation

The line containing the reflector, $\mathbf{l}_{AB}$, can be estimated using the estimated image source position $\hat{\boldsymbol{x}}_{S'}$ and the source position $\hat{\boldsymbol{x}}_S$. In fact, the line on which the reflector $\boldsymbol{x}_A \boldsymbol{x}_B$ lies $\mathbf{l}_{AB}$ is the axis of the segment $\boldsymbol{x}_S \boldsymbol{x}_{S'}$. Therefore $\mathbf{l}_{AB}$ is found by imposing that $\boldsymbol{x}_{S'}$ is a mirrored version of $\boldsymbol{x}_S$ across the reflector line $\mathbf{l}_{AB}$ as illustrated in Figure 5.3.

Given $\hat{\mathbf{x}}_{S'}$ and $\hat{\mathbf{x}}_S$ (homogeneous coordinates of the image source and source, respectively) and using (2.3), the line joining the source and the image source is given by $\mathbf{l}_{SS'} = [l_{1SS'}, l_{2SS'}, l_{3SS'}]^T = \hat{\mathbf{x}}_S \times \hat{\mathbf{x}}_{S'}$, while the line $\mathbf{l}' = [l'_1, l'_2, l'_3]^T = \hat{\mathbf{x}}_S - \hat{\mathbf{x}}_{S'}$ is the line passing through the origin ($l'_3 = 0$) perpendicular to $\mathbf{l}_{SS'}$ as it has the negative reciprocal of the slope of $\mathbf{l}_{SS'}$, i.e. $l'_1/l'_2 = -l_{2SS'}/l_{1SS'}$. The reflector line $\mathbf{l}_{AB} = [l_{1AB}, l_{2AB}, l_{3AB}]^T$ has the same direction as $\mathbf{l}'$ but it passes through the middle point $\mathbf{x}_m = \frac{1}{2}(\hat{\mathbf{x}}_S + \hat{\mathbf{x}}_{S'})$, i.e. $\mathbf{x}_m^T \mathbf{l}_{AB} = 0$. Given $[l_{1AB}, l_{2AB}]^T = \hat{\boldsymbol{x}}_S - \hat{\boldsymbol{x}}_{S'}$ we have to find the third coordinate l_{3AB} such that $\mathbf{x}_m^T \mathbf{l}_{AB} = \frac{1}{2}(\hat{\boldsymbol{x}}_S + \hat{\boldsymbol{x}}_{S'})^T(\hat{\boldsymbol{x}}_S - \hat{\boldsymbol{x}}_{S'}) + l_{3AB} = 0$. Finally we get

$$\hat{\mathbf{l}}_{AB} = \begin{bmatrix} \hat{\boldsymbol{x}}_S - \hat{\boldsymbol{x}}_{S'} \\ \frac{1}{2}(\hat{\boldsymbol{x}}_{S'}^T \hat{\boldsymbol{x}}_{S'} - \hat{\boldsymbol{x}}_S^T \hat{\boldsymbol{x}}_S) \end{bmatrix}. \tag{5.7}$$

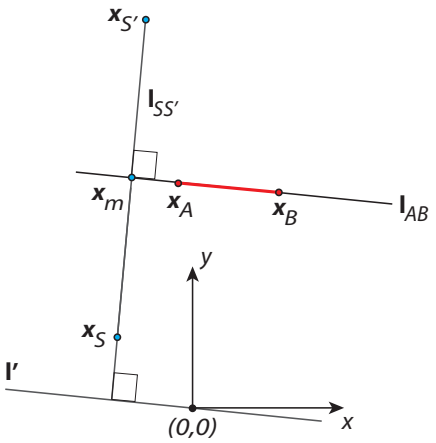


Figure 5.3.: Reflector line estimation.

5.6. Reflector endpoints estimation

Consider now the plenacoustic image $\mathcal{I}(\mathbf{l})$ that captures the reflection originated from the image source $\boldsymbol{x}_{S'}$ generated by the reflector $\boldsymbol{x}_A\boldsymbol{x}_B$. For the sake of clarity in the illustration, we assume in this example that only the image source is visible in $\mathcal{I}(\mathbf{l})$. As observed before, not always all the microphones sense the image source $\boldsymbol{x}_{S'}$. It may happen that the image source is visible only from a limited set of microphones in the array, due to the fact that reflective acoustic paths are bound to pass through the segment $\boldsymbol{x}_A\boldsymbol{x}_B$, as depicted in Figure 5.2 (a). In the ray space the visibility of $\overline{\mathbf{x}}_{S'}$ is limited by the dimensions of the reflector, as depicted in Figure 5.2 (b). The edges of the reflector visibility region, $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$, are related to the endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. As a consequence, observing the edges of the visibility region in the ray space allows the estimation of the reflector's endpoints in the geometric space.

5.6.1. Edge rays

To be observable the edges $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$ should be "illuminated" by the image source $\overline{\mathbf{x}}_{S'}$. By knowing the image source position and the last microphone(s) that sense the reflective paths originating from the image source we can estimate the *edge ray*(s) that lie on the edge of the reflector's visibility region, $\mathbf{l}_A = [l_{1A}, l_{2A}, l_{3A}]^T$ and/or $\mathbf{l}_B = [l_{1B}, l_{2B}, l_{3B}]^T$. In the following analysis we consider specular reflection only, for the discussion of a case in which diffraction becomes significant see the considerations 5.10 at the end of the section.

We start finding the point of intersection between the edge ray $\mathbf{l}_A$ and the microphone array. Using (5.1) we find the last microphone(s) that sense the reflective paths originating from the image source. Let us suppose that $\boldsymbol{x}_e$ is the last microphone that does not sense the image source, and $\boldsymbol{x}_{e+1}$ the first one that does (see Figure 5.2 (a)). Due to the finite number of microphones, we cannot estimate the exact point where the edge ray $\mathbf{l}_A$ intersects the array. However we know that this intersection point is between $\boldsymbol{x}_e$ and $\boldsymbol{x}_{e+1}$. We expect that in average it intersects the array in the middle point $\boldsymbol{x}_t = (\boldsymbol{x}_e + \boldsymbol{x}_{e+1})/2$ and, therefore, we use $\boldsymbol{x}_t$ as the best estimate of the intersection point between the edge ray $\mathbf{l}_A$ and the microphone array.

Given the intersection point estimate $\boldsymbol{x}_t$ and the image source position estimate $\hat{\boldsymbol{x}}_{S'}$, we estimate the edge ray $\mathbf{l}_A$ as the line joining $\boldsymbol{x}_t$ and $\hat{\boldsymbol{x}}_{S'}$ in the geometric space (see Figure 5.2 (a)) or, equivalently, as the intersection of lines $\overline{\mathbf{x}}_t$ and $\overline{\mathbf{x}}_{S'}$ in the reduced ray space (Figure 5.2 (b)). Therefore, we can write

$$\begin{aligned} \hat{\mathbf{x}}_{S'}^T \mathbf{l}_A &= 0, \\ \mathbf{x}_t^T \mathbf{l}_A &= 0. \end{aligned} \tag{5.8}$$

Notice from (5.8) that $\mathbf{l}_A$ is a vector spanning the null space of $[\hat{\mathbf{x}}_{S'}, \mathbf{x}_t]^T$. Under the assumption that $l_{3A} \neq 0$, we can normalize it to 1 (recall that line parameters are scalable). Indicating $\hat{\mathbf{G}} = [\hat{\boldsymbol{x}}_{S'}, \boldsymbol{x}_t]^T$ and $\mathbf{1} = [1, 1]^T$, we can write

$$\hat{\mathbf{G}} \mathbf{l}_A^* = -\mathbf{1}$$

and therefore $\mathbf{l}_A^* = [l_{1A}^*, l_{2A}^*]^T = [l_{1A}/l_{3A}, l_{2A}/l_{3A}]^T$ is given by

$$\hat{\mathbf{l}}_A^* = -\hat{\mathbf{G}}^{-1} \mathbf{1}. \tag{5.9}$$

The edge ray estimate is finally given by $\hat{\mathbf{l}}_A = [\hat{\mathbf{l}}_A^{*T}, 1]^T$.

5.6.2. Endpoints

A single edge ray is not sufficient for the estimation of the reflector endpoints. We can easily visualize this fact in the ray space. In fact, the reflector endpoint corresponds to a line in the reduced ray space and the found edge ray lies on this line, as it originates from the endpoint (see Figure 5.2). Therefore, in order to

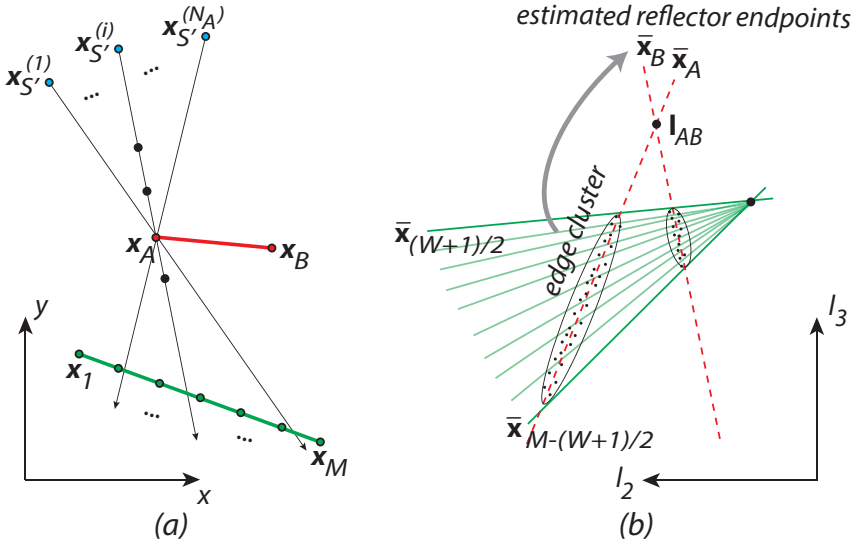


Figure 5.4.: (a) Configuration in the geometric space; (b) Configuration in the reduced ray space: linear regression is used to estimate the reflector endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$; intersection between $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$ represents the reflector line $\mathbf{l}_{AB}$, i.e. the line joining the endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$.

infer a line from points lying on it we need at least two points, i.e. two edge rays. In the geometric space this means that different endpoint positions can cause the same visibility discontinuity (edge ray) to be observable by the microphone array as evidenced by the black dots in Figure 5.4 (a). For this reason, the estimation of the endpoints of the reflector is possible when at least a couple of plenacoustic images, acquired for different source locations, are available, as shown in Figure 5.4 (b).

Moving $\boldsymbol{x}_S$ at different locations we acquire a number of different plenacoustic images and as a consequence a number of edge rays $\hat{\mathbf{l}}_A^{(i)}$, $i = 1, ..., N_A$ and $\hat{\mathbf{l}}_B^{(j)}$, $j = 1, ..., N_B$ (see Figure 5.4 (b)), which lie on the edges of the visibility region, i.e. on the lines $\overline{\mathbf{x}}_A$ and $\overline{\mathbf{x}}_B$. By fitting lines to edge rays we estimate reflector endpoints $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$. In a similar way as done for the source position in (5.3), (5.4), (5.5) and (5.6), we estimate the reflector endpoint positions using the LS method

$$\begin{aligned} \hat{\boldsymbol{x}}_A &= (\hat{\mathbf{H}}_A^T \hat{\mathbf{H}}_A)^{-1} \hat{\mathbf{H}}_A^T \hat{\mathbf{d}}_A, \\ \hat{\boldsymbol{x}}_B &= (\hat{\mathbf{H}}_B^T \hat{\mathbf{H}}_B)^{-1} \hat{\mathbf{H}}_B^T \hat{\mathbf{d}}_B, \end{aligned} \tag{5.10}$$

where the matrices $\hat{\mathbf{H}}_A$, $\hat{\mathbf{H}}_B$ and the vectors $\hat{\mathbf{d}}_A$, $\hat{\mathbf{d}}_B$ are formed analogously to $\hat{\mathbf{H}}_{S'}$ and $\hat{\mathbf{d}}_{S'}$, respectively.

As a consequence, the endpoint estimation is, once more, a linear problem in the

ray space. Fitting lines to edge rays in the reduced ray space corresponds in the geometric space to finding the intersection of the given edge rays, i.e. finding the endpoint position as shown in Figure 5.4 (a).

5.7. Performance analysis

In order to analyze the accuracy of source and reflector endpoint estimators we derive expressions of second order statistics of estimates in (5.2), (5.6), (5.9) and (5.10). We start considering the accuracy of (5.2). The points $\hat{\mathbf{l}}_{S'}^{(i)}$ correspond to maxima of $P_i(\mathbf{l})$, i.e. to the maximum of $P_i(\theta)$. The maximum of $P_i(\theta)$ is the estimated direction of arrival $\hat{\theta}_i = \theta_i + \epsilon_i$, where ϵ_i is the estimation error. Given the DOA estimation technique we suppose that we know the $N_{S'} \times N_{S'}$ covariance matrix of the $N_{S'}$ estimates $\hat{\theta}_i$, $\mathbf{C}_\theta$, which depends on the the beamforming method; the number of microphones; their spacing; the signal frequency; direction of arrival; and the signal-to-noise ratio (SNR). The statistical performance for the most used bemforming techniques can be found in [91] or alternatively estimated from experiments/simulations. We suppose DOA estimates to be unbiased, i.e. ϵ_i to be zero averaged.

Supposing small errors on DOA estimates we can expand $\hat{\mathbf{l}}_{S'}^{(i)} = \mathbf{l}_{S'}^{(i)}(\hat{\theta}_i)$ in the first-order Taylor series around the true value $\mathbf{l}_{S'}^{(i)} = \mathbf{l}_{S'}^{(i)}(\theta_i)$, i.e.

$$\mathbf{l}_{S'}^{(i)}(\hat{\theta}_i) \simeq \mathbf{l}_{S'}^{(i)}(\theta_i) + \frac{\partial \mathbf{l}_{S'}^{(i)}(\theta_i)}{\partial \theta_i}^T (\hat{\theta}_i - \theta_i). \tag{5.11}$$

Using (5.11) we approximate the covariances of estimated points as

$$\begin{aligned}
\left[\mathbf{C}_{\mathbf{l}_{S'}}\right]_{ij} &= E\left[(\mathbf{l}_{S'}^{(i)}(\hat{\theta}_i) - E[\mathbf{l}_{S'}^{(i)}(\hat{\theta}_i)])(\mathbf{l}_{S'}^{(j)}(\hat{\theta}_j) - E[\mathbf{l}_{S'}^{(j)}(\hat{\theta}_j)])^T\right] \\
&\simeq E\left[(\mathbf{l}_{S'}^{(i)}(\hat{\theta}_i) - \mathbf{l}_{S'}^{(i)}(\theta_i))(\mathbf{l}_{S'}^{(j)}(\hat{\theta}_j) - \mathbf{l}_{S'}^{(j)}(\theta_j))^T\right] \\
&\simeq \frac{\partial \mathbf{l}_{S'}^{(i)}(\theta_i)}{\partial \theta_i}^T E\left[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\right] \frac{\partial \mathbf{l}_{S'}^{(j)}(\theta_j)}{\partial \theta_j} \\
&= \frac{\partial \mathbf{l}_{S'}^{(i)}(\theta_i)}{\partial \theta_i}^T [\mathbf{C}_\theta]_{ij} \frac{\partial \mathbf{l}_{S'}^{(j)}(\theta_j)}{\partial \theta_j},
\end{aligned} \tag{5.12}$$

where $E[\mathbf{.}]$ is the expectation operator; $[\mathbf{C}_\theta]_{ij}$ indicates the covariance between DOA estimates $\hat{\theta}_i$ and $\hat{\theta}_j$; $\mathbf{C}_{\mathbf{l}_{S'}}$ is the $3N_{S'} \times 3N_{S'}$ covariance matrix of estimated rays and $\left[\mathbf{C}_{\mathbf{l}_{S'}}\right]_{ij}$ indicates the 3×3 submatrix containing the covariances between coefficients

of rays $\hat{\mathbf{l}}_{S'}^{(i)}$ and $\hat{\mathbf{l}}_{S'}^{(j)}$. From (4.5) with $k = 1$ and, for simplicity, $\alpha = \pi/2$ we obtain

$$\frac{\partial \mathbf{l}_{S'}^{(i)}(\theta_i)}{\partial \theta_i}^T = [\sin(\theta_i), \cos(\theta_i), -\sin(\theta_i)x_i - \cos(\theta_i)y_i]^T . \tag{5.13}$$

If we replace (5.13) in (5.12) we get the desired covariance matrix.

We now analyze the accuracy of the LS estimation of the image source position given by (5.6). The estimates $\hat{\mathbf{l}}_{S'}^{(i)}$ are affected by estimation error $\mathbf{l}_{S'}^{(\varepsilon i)}$, whose covariance matrix is given by (5.12), i.e. $\hat{\mathbf{l}}_{S'}^{(i)} = \mathbf{l}_{S'}^{(i)} + \mathbf{l}_{S'}^{(\varepsilon i)}$, where $\mathbf{l}_{S'}^{(i)}$ is the true value of the given ray. As a consequence, in (5.4) the matrix $\hat{\mathbf{M}}_{S'}$ is affected by error $\mathbf{M}_{S'}^{(\varepsilon)} = [\hat{\mathbf{l}}_{S'}^{(\varepsilon i_1)}, \hat{\mathbf{l}}_{S'}^{(\varepsilon i_2)}, ..., \hat{\mathbf{l}}_{S'}^{(\varepsilon i_{N_{S'}})}]^T$, $i_1, i_2, ..., i_{N_{S'}} \in \mathcal{V}_{S'}$, i.e. $\hat{\mathbf{M}}_{S'} = \mathbf{M}_{S'} + \mathbf{M}_{S'}^{(\varepsilon)}$, where $\mathbf{M}_{S'}$ is the matrix that contains the true values $\mathbf{l}_{S'}^{(i)}$, $i \in \mathcal{V}_{S'}$. Therefore, writing $\mathbf{M}_{S'} = [\mathbf{H}_{S'}, -\mathbf{d}_{S'}]$ and $\mathbf{M}_{S'}^{(\varepsilon)} = [\mathbf{H}_{S'}^{(\varepsilon)}, -\mathbf{d}_{S'}^{(\varepsilon)}]$, in (5.5) both $\hat{\mathbf{H}}_{S'}$ and $\hat{\mathbf{d}}_{S'}$ are affected by estimation error, i.e. $\hat{\mathbf{H}}_{S'} = \mathbf{H}_{S'} + \mathbf{H}_{S'}^{(\varepsilon)}$ and $\hat{\mathbf{d}}_{S'} = \mathbf{d}_{S'} + \mathbf{d}_{S'}^{(\varepsilon)}$.

Notice that, given the presence of noise in the data matrix $\hat{\mathbf{H}}_{S'}$ and not just in the observations $\hat{\mathbf{d}}_{S'}$, the ordinary least-squares (5.6) is in general biased. For the derivation of the variance of estimates we assume this bias to be negligible. We model both errors as additive noises

$$\begin{aligned}
\hat{\mathbf{H}}_{S'}\boldsymbol{x}_{S'} &= \hat{\mathbf{d}}_{S'} \\
(\mathbf{H}_{S'} + \mathbf{H}_{S'}^{(\varepsilon)})\boldsymbol{x}_{S'} &= \mathbf{d}_{S'} + \mathbf{d}_{S'}^{(\varepsilon)} \\
\mathbf{H}_{S'}\boldsymbol{x}_{S'} &= \mathbf{d}_{S'} - \mathbf{H}_{S'}^{(\varepsilon)}\boldsymbol{x}_{S'} + \mathbf{d}_{S'}^{(\varepsilon)},
\end{aligned}$$

and write

$$\mathbf{H}_{S'}\boldsymbol{x}_{S'} = \mathbf{d}_{S'} + \mathbf{n}_{S'},$$

where the noise term $\mathbf{n}_{S'}$ is given by

$$\mathbf{n}_{S'} = -\mathbf{H}_{S'}^{(\varepsilon)}\boldsymbol{x}_{S'} + \mathbf{d}_{S'}^{(\varepsilon)} = -\mathbf{M}_{S'}^{(\varepsilon)}\mathbf{x}_{S'}.$$

Given the covariance matrix in (5.12), we compute the covariance matrix of $\mathbf{n}_{S'}$ as

$$\begin{aligned}
\left[\mathbf{C}_{\mathbf{n}_{S'}}\right]_{ij} \simeq E\left[\mathbf{n}_{S'}\mathbf{n}_{S'}^T\right]_{ij} &= E\left[(-\mathbf{x}_{S'}^T\mathbf{l}_A^{(\varepsilon i)})(-\mathbf{l}_A^{(\varepsilon j)T}\mathbf{x}_{S'})\right] = \mathbf{x}_{S'}^T\left[\mathbf{C}_{\mathbf{l}_{S'}}\right]_{ij}\mathbf{x}_{S'} \\
&= [\sin(\theta_i), \cos(\theta_i)](\boldsymbol{x}_{S'} - \boldsymbol{x}_i)\left[\mathbf{C}_\theta\right]_{ij}[\sin(\theta_j), \cos(\theta_j)](\boldsymbol{x}_{S'} - \boldsymbol{x}_j).
\end{aligned} \tag{5.14}$$

The covariance matrix of LS estimate $\hat{\boldsymbol{x}}_{S'}$ in (5.6) is then approximated by

$$\begin{aligned}\mathbf{C}_{\boldsymbol{x}_{S'}} = E\left[(\hat{\boldsymbol{x}}_{S'} - E[\hat{\boldsymbol{x}}_{S'}])(\hat{\boldsymbol{x}}_{S'} - E[\hat{\boldsymbol{x}}_{S'}])^T\right] &\simeq E\left[(\hat{\boldsymbol{x}}_{S'} - \boldsymbol{x}_{S'})(\hat{\boldsymbol{x}}_{S'} - \boldsymbol{x}_{S'})^T\right] \\ &\simeq (\mathbf{H}_{S'}^T\mathbf{H}_{S'})^{-1}\mathbf{H}_{S'}^T E\left[\mathbf{n}_{S'}\mathbf{n}_{S'}^T\right]\mathbf{H}_{S'}(\mathbf{H}_{S'}^T\mathbf{H}_{S'})^{-1} \\ &= (\mathbf{H}_{S'}^T\mathbf{H}_{S'})^{-1}\mathbf{H}_{S'}^T\mathbf{C}_{\mathbf{n}_{S'}}\mathbf{H}_{S'}(\mathbf{H}_{S'}^T\mathbf{H}_{S'})^{-1}.\end{aligned} \tag{5.15}$$

The estimate in (5.9) is affected by two independent kinds of error:

1. Error $\boldsymbol{x}_t^{(\varepsilon)}$ on the intersection point estimate $\boldsymbol{x}_t = \boldsymbol{x}_{int} + \boldsymbol{x}_t^{(\varepsilon)}$ ($\boldsymbol{x}_{int}$ being the true intersection point) caused by the sampling of the array. This error depends on geometry and is constant for given source and array positions. As a consequence it adds a bias to the estimation of $\mathbf{l}_A^*$ that can be controlled decreasing the spacing between microphones;

2. Error $\boldsymbol{x}_{S'}^{(\varepsilon)}$ on the image source position estimate $\hat{\boldsymbol{x}}_{S'} = \boldsymbol{x}_{S'} + \boldsymbol{x}_{S'}^{(\varepsilon)}$ with covariance given by (5.15). This error is caused by the discrete nature of the array, which causes blurring to appear in the plenacoustic image.

Notice that $\hat{\mathbf{G}} = \mathbf{G} + [\boldsymbol{x}_{S'}^{(\varepsilon)}, \boldsymbol{x}_t^{(\varepsilon)}]^T$; equation (5.9) becomes $\mathbf{G}\mathbf{l}_A^* = -\mathbf{1} - [\boldsymbol{x}_{S'}^{(\varepsilon)}, \boldsymbol{x}_t^{(\varepsilon)}]^T\mathbf{l}_A^*$ and, supposing the bias to be negligible, the variance of the estimate $\mathbf{l}_A^*$ is approximated as

$$\begin{aligned}&var(\mathbf{l}_A^*) \simeq E\left[(\hat{\mathbf{l}}_A^* - \mathbf{l}_A^*)(\hat{\mathbf{l}}_A^* - \mathbf{l}_A^*)^T\right] \\ &\simeq E\left[\left(\mathbf{G}^{-1}(-\mathbf{1} - \begin{bmatrix}\boldsymbol{x}_{S'}^{(\varepsilon)T} \\ \boldsymbol{x}_t^{(\varepsilon)T}\end{bmatrix}\mathbf{l}_A^*) + \mathbf{G}^{-1}\mathbf{1}\right)\left(\mathbf{G}^{-1}(-\mathbf{1} - \begin{bmatrix}\boldsymbol{x}_{S'}^{(\varepsilon)T} \\ \boldsymbol{x}_t^{(\varepsilon)T}\end{bmatrix}\mathbf{l}_A^*) + \mathbf{G}^{-1}\mathbf{1}\right)^T\right] \\ &= \mathbf{G}^{-1}E\left[\begin{bmatrix}\boldsymbol{x}_{S'}^{(\varepsilon)T}\mathbf{l}_A^* \\ \boldsymbol{x}_t^{(\varepsilon)T}\mathbf{l}_A^*\end{bmatrix}\left[\mathbf{l}_A^{*T}\boldsymbol{x}_{S'}^{(\varepsilon)}, \mathbf{l}_A^{*T}\boldsymbol{x}_t^{(\varepsilon)}\right]\right]\mathbf{G}^{-T} = \mathbf{G}^{-1}\begin{bmatrix}c_{S'} & 0 \\ 0 & c_t\end{bmatrix}\mathbf{G}^{-T},\end{aligned} \tag{5.16}$$

where $E[(\mathbf{l}_A^{*T}\boldsymbol{x}_{S'}^{(\varepsilon)})(\boldsymbol{x}_t^{(\varepsilon)T}\mathbf{l}_A^*)] = 0$; given the covariance matrix (5.15),

$$c_{S'} \simeq E\left[(\mathbf{l}_A^{*T}\boldsymbol{x}_{S'}^{(\varepsilon)})(\boldsymbol{x}_{S'}^{(\varepsilon)T}\mathbf{l}_A^*)\right] = \mathbf{l}_A^{*T}E\left[\boldsymbol{x}_{S'}^{(\varepsilon)}\boldsymbol{x}_{S'}^{(\varepsilon)T}\right]\mathbf{l}_A^* = \mathbf{l}_A^{*T}\mathbf{C}_{\boldsymbol{x}_{S'}}\mathbf{l}_A^*,$$

and $c_t \simeq E[(\mathbf{l}_A^{*T}\boldsymbol{x}_t^{(\varepsilon)})(\boldsymbol{x}_t^{(\varepsilon)T}\mathbf{l}_A^*)] = 0$. For the multiple edge ray estimates $\hat{\mathbf{l}}_A^{*(i)}$, $i = 1, ..., N_A$ we have the covariance matrix $\mathbf{C}_{\mathbf{l}_A^*}$ with $\left[\mathbf{C}_{\mathbf{l}_A^*}\right]_{ii} = var(\mathbf{l}_A^{*(i)})$ and $\left[\mathbf{C}_{\mathbf{l}_A^*}\right]_{ij} = \mathbf{0}$ for $i \neq j$ as different edge rays, $\mathbf{l}_A^{*(i)}$ and $\mathbf{l}_A^{*(j)}$, are obtained with differ-

ent observations, i.e. moving the source, and are therefore uncorrelated.

Finally the covariances of estimates in (5.10) are obtained similarly to (5.15)

$$\begin{aligned}\mathbf{C}_{\boldsymbol{x}_A} &= (\mathbf{H}_A^T\mathbf{H}_A)^{-1}\mathbf{H}_A^T\mathbf{C}_{\mathbf{n}_A}\mathbf{H}_A(\mathbf{H}_A^T\mathbf{H}_A)^{-1},\\ \mathbf{C}_{\boldsymbol{x}_B} &= (\mathbf{H}_B^T\mathbf{H}_B)^{-1}\mathbf{H}_B^T\mathbf{C}_{\mathbf{n}_B}\mathbf{H}_B(\mathbf{H}_B^T\mathbf{H}_B)^{-1},\end{aligned} \tag{5.17}$$

where, given (5.16),

$$\begin{aligned}[\mathbf{C}_{\mathbf{n}_A}]_{ii} &= \boldsymbol{x}_A^T\left[\mathbf{C}_{\mathbf{l}_A^*}\right]_{ii}\boldsymbol{x}_A,\\ [\mathbf{C}_{\mathbf{n}_B}]_{ii} &= \boldsymbol{x}_B^T\left[\mathbf{C}_{\mathbf{l}_B^*}\right]_{ii}\boldsymbol{x}_B.\end{aligned} \tag{5.18}$$

The given analysis can be used to assess the performance of the system without the need of running a set of time-consuming Monte Carlo simulations. Furthermore the theoretical covariance matrices can be used to improve the estimation accuracy as detailed in the nex section.

5.8. Iterative WLS estimation

The previous performance analysis is useful not only for system design but can also be used to improve the estimation accuracy. In fact, if we apply the weighted least-squares (WLS) estimators in (5.6) and (5.10) become, respectively

$$\begin{aligned}\hat{\boldsymbol{x}}_{S'} &= (\hat{\mathbf{H}}_{S'}^T\mathbf{W}_{S'}\hat{\mathbf{H}}_{S'})^{-1}\hat{\mathbf{H}}_{S'}^T\mathbf{W}_{S'}\hat{\mathbf{d}}_A,\\ \hat{\boldsymbol{x}}_A &= (\hat{\mathbf{H}}_A^T\mathbf{W}_A\hat{\mathbf{H}}_A)^{-1}\hat{\mathbf{H}}_A^T\mathbf{W}_A\hat{\mathbf{d}}_A,\\ \hat{\boldsymbol{x}}_B &= (\hat{\mathbf{H}}_B^T\mathbf{W}_B\hat{\mathbf{H}}_B)^{-1}\hat{\mathbf{H}}_B^T\mathbf{W}_B\hat{\mathbf{d}}_B,\end{aligned} \tag{5.19}$$

and, as a consequence, the covariance matrices in (5.15) and (5.17) are, respectively

$$\begin{aligned}\mathbf{C}_{\boldsymbol{x}_{S'}} &= (\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{H}_{S'})^{-1}\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{C}_{\mathbf{n}_{S'}}\mathbf{W}_{S'}^T\mathbf{H}_{S'}(\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{H}_{S'})^{-1},\\ \mathbf{C}_{\boldsymbol{x}_A} &= (\mathbf{H}_A^T\mathbf{W}_A\mathbf{H}_A)^{-1}\mathbf{H}_A^T\mathbf{W}_A\mathbf{C}_{\mathbf{n}_A}\mathbf{W}_A^T\mathbf{H}_A(\mathbf{H}_A^T\mathbf{W}_A\mathbf{H}_A)^{-1},\\ \mathbf{C}_{\boldsymbol{x}_B} &= (\mathbf{H}_B^T\mathbf{W}_B\mathbf{H}_B)^{-1}\mathbf{H}_B^T\mathbf{W}_B\mathbf{C}_{\mathbf{n}_B}\mathbf{W}_B^T\mathbf{H}_B(\mathbf{H}_B^T\mathbf{W}_B\mathbf{H}_B)^{-1}.\end{aligned}$$

If $\mathbf{W}_{S'} = \mathbf{C}_{\mathbf{n}_{S'}}^{-1}$, $\mathbf{W}_A = \mathbf{C}_{\mathbf{n}_A}^{-1}$ and $\mathbf{W}_B = \mathbf{C}_{\mathbf{n}_B}^{-1}$ we obtain the maximum likelihood (ML) [92] estimates of $\hat{\boldsymbol{x}}_{S'}$, $\hat{\boldsymbol{x}}_A$ and $\hat{\boldsymbol{x}}_B$, with

$$\mathbf{C}_{\boldsymbol{x}_{S'}} = (\mathbf{H}_{S'}^T\mathbf{C}_{\mathbf{n}_{S'}}^{-1}\mathbf{H}_{S'})^{-1},\ \mathbf{C}_{\boldsymbol{x}_A} = (\mathbf{H}_A^T\mathbf{C}_{\mathbf{n}_A}^{-1}\mathbf{H}_A)^{-1},\ \mathbf{C}_{\boldsymbol{x}_B} = (\mathbf{H}_B^T\mathbf{C}_{\mathbf{n}_B}^{-1}\mathbf{H}_B)^{-1}.$$

However covariance matrices $\mathbf{C}_{\mathbf{n}_{S'}}$, $\mathbf{C}_{\mathbf{n}_A}$ and $\mathbf{C}_{\mathbf{n}_B}$ depend on source and reflector positions (see (5.14) and (5.18)) and therefore they cannot be known a priori. Nonetheless we propose an iterative algorithm that approaches the ML estimator. The algorithm works as follows:

1. obtain initial estimates $\hat{\boldsymbol{x}}_{S'}$, $\hat{\boldsymbol{x}}_A$ and $\hat{\boldsymbol{x}}_B$ with $\mathbf{W}_{S'} = \mathbf{I}$, $\mathbf{W}_A = \mathbf{I}$ and $\mathbf{W}_B = \mathbf{I}$;
2. compute $\hat{\mathbf{C}}_{\mathbf{n}_{S'}}$, $\hat{\mathbf{C}}_{\mathbf{n}_A}$ and $\hat{\mathbf{C}}_{\mathbf{n}_B}$ for the estimated positions;
3. obtain new estimates $\hat{\boldsymbol{x}}_{S'}$, $\hat{\boldsymbol{x}}_A$ and $\hat{\boldsymbol{x}}_B$ with $\mathbf{W}_{S'} = \hat{\mathbf{C}}^{-1}_{\mathbf{n}_{S'}}$, $\mathbf{W}_A = \hat{\mathbf{C}}^{-1}_{\mathbf{n}_A}$ and $\mathbf{W}_B = \hat{\mathbf{C}}^{-1}_{\mathbf{n}_B}$;
4. repeat from step 2.

The iterative procedure stops when a prescribed number of iterations is reached. The simulations presented at the end of this Chapter, and in particular the simulation 3 for reflector localization, show that the algorithm gets close to the ML performance in just one iteration, with a significant gain with respect to the LS estimates.

5.9. Multiple sources and reflectors

In this paragraph we extend the detection method to the case of multiple sources and multiple reflectors. Some modifications are necessary in order to account for the presence of multiple "objects" in the acoustic scene. For this extension the use of the ray space becomes very useful.

On the plenacoustic image the contributions of different objects can be fairly easily distinguished and extracted. Consider the case when multiple sources (direct and image, for example) are present. The pseudospectra acquired by each subarray presents a number of peaks corresponding to Directions Of Arrival of different acoustic sources. Therefore, a labeling problem arises, as we need to keep track across all the pseudospectra of which DOA corresponds to which source. On the other hand, if measurements are represented as plenacoustic images in the reduced ray space, they naturally cluster around the searched source positions. In fact, points corresponding to rays coming from the same source should all stay on the same line, i.e. on the ray space representation of the given source. This is true also for the clustering of edge rays corresponding to different endpoints as they are represented as lines on the image as well. Without the ray space representation keeping track of which edge ray corresponds to which endpoint would be a more difficult problem.

The estimation procedure in the case of multiple sources and reflectors is summarized in Figure 5.5. First, when multiple image sources are present, we need to distinguish among different sources. We do this through the Hough transform. We then check the presence of edge rays for all the estimated sources. The estimated edge rays are then clustered in turn, each one related to the different endpoint. By fitting lines to the points of each cluster we estimate the (visible) endpoints of all reflectors in the environment. Details of each step follow.

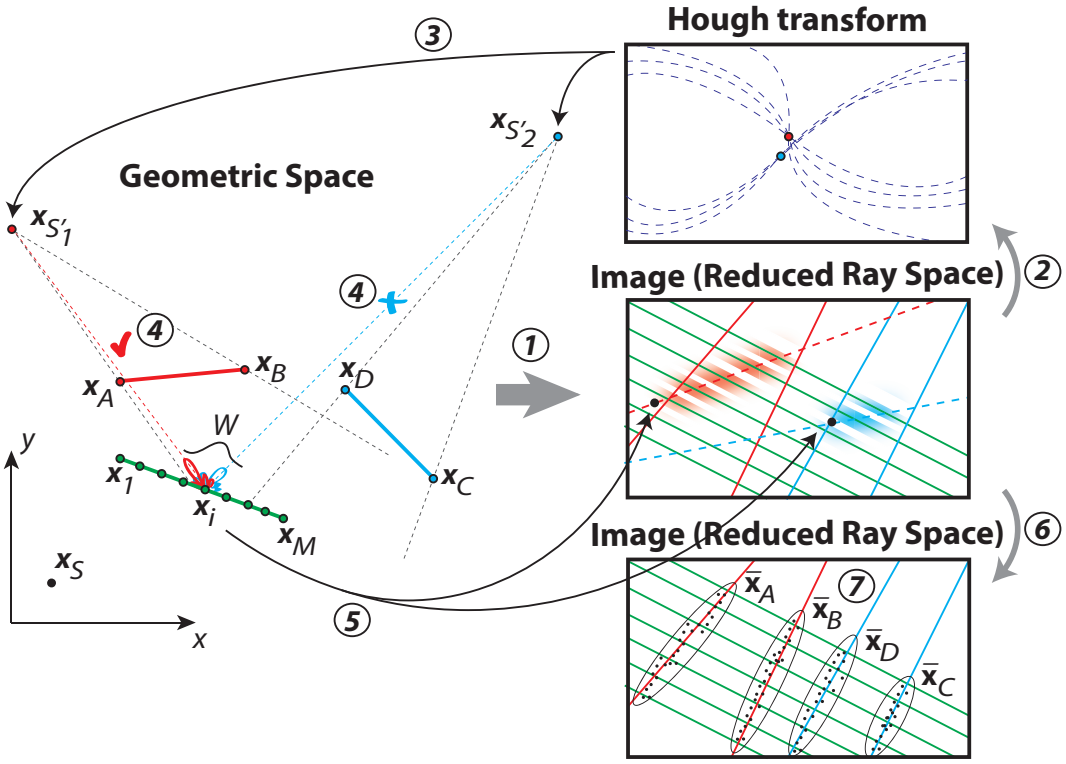


Figure 5.5.: Estimation steps: 1) acoustic image is obtained from acoustic measurements; 2) Hough transform is used in order to find lines in the image; 3) positions of the image sources are estimated; 4) for each position the beamforming techniques are used in order to test the visibility of the image sources; 5) if observable the edge rays are found and mapped to the ray space; 6) a number of edge rays are obtained moving the acoustic source; 7) clusters of edge rays are used to estimate reflector endpoints.

5.9.1. Localization of multiple sources

Let us consider the problem of localizing multiple acoustic sources. The first step is the disambiguation of measurements (DOAs, TOAs, TDOAs) obtained from the arrays and their matching to the corresponding sources. Disambiguation of

TDOAs, for example, can be performed as in [93]. A method for matching measurements and sources is proposed in [94], based on a Guassian likelihood function. When using plenacoustic imaging, the disambiguation and pairing of information is greatly simplified because sources are imaged as well-distinguished linear features. Disambiguation can be readily accomplished using a Hough transform ([95], [96]) on the plenacoustic image. A detailed description follows.

In presence of multiple sources $P_i(\mathbf{l})$ exhibits multiple peaks, one for each visible source. We select these peaks by finding local maxima in $P_i(\mathbf{l})$ that overcome the threshold $\beta \max(\mathcal{I}(\mathbf{l}))$:

$$\mathcal{L}_i = \{\mathbf{l} \,|P_i(\mathbf{l}-\Delta\mathbf{l}) < P_i(\mathbf{l}) > P_i(\mathbf{l}+\Delta\mathbf{l}),$$
$$P_i(\mathbf{l}) > \beta \max(\mathcal{I}(\mathbf{l}))\}.$$

where $\Delta\mathbf{l}$ defines a neighborhood in which local maxima are found. By finding the local maxima of all $P_i(\mathbf{l})$ we obtain the set of peaks on the image $\mathcal{P} = \{\mathcal{L}_i \mid i = (W+1)/2, \ldots, M-(W+1)/2\}$. This set contains rays (peaks) generated by different acoustic sources $\boldsymbol{x}_{S_j}$, $j = 1, ..., N_r$, where N_r is the number of sources. Each acoustic source $\boldsymbol{x}_{S_j}$ corresponds in the reduced ray space to the line $\overline{\mathbf{x}}_{S_j}$. As a consequence the rays in $\mathcal{P}$ generated by different sources should stay on different lines. Starting from the set $\mathcal{P}$, these lines can be found using the Hough transform [96], [97], [98]. The Hough transform parametrizes a straight line in two dimensions as

$$\varrho = x\cos(\theta) + y\sin(\theta), \tag{5.20}$$

where θ is the angle of the line's normal; ϱ is the distance between the line and the origin and x and y are the coordinates of the point on the image that, in our case, correspond to the coefficients of the reduced ray space used to represent the plenacoustic image. In particular, in simulations and experiments we obtain the reduced ray space using the cutting plane $l_1 = 1$[1] and, as a consequence, we have $x = l_2$, $y = l_3$. The Hough transform (5.20) becomes

$$\varrho = l_2\cos(\theta) + l_3\sin(\theta). \tag{5.21}$$

[1] Recall that this particular choice of reduced ray space is fully equivalent to the projective ray space except for the rays that are parallel to the x-axis, which we cannot represent because they have $l_1 = 0$.

A point (ray) in the plenacoustic image maps to a sinusoidal curve in the Hough space. This sinusoid corresponds to all the lines passing through the given point on the plenacoustic image, which in turn correspond to all the points that the ray encounters in the geometric space. Points in the Hough space correspond to the lines on the plenacoustic image, i.e. to the points in the geometric space. The rays $\mathcal{P}$ that lie on the same line $\overline{\mathbf{x}}_{S_j}$ (source $\boldsymbol{x}_{S_j}$) in the plenacoustic image map in the Hough space to sinusoids that intersect in the same point (ϱ_j, θ_j). The point (ϱ_j, θ_j) (Hough space) parametrizes the line $\overline{\mathbf{x}}_{S_j}$ (plenacoustic image) and, consequently, the source position $\boldsymbol{x}_{S_j}$ (geometric space).

In order to estimate the positions of different sources the peaks $\mathcal{P}$ are clustered using the Hough transform as follows. First the accumulator $\mathcal{A}(\varrho, \theta)$ is initialized as $\mathcal{A}(\varrho, \theta) = 0,\ \varrho = 0,\ \Delta\varrho,\ 2\Delta\varrho, ...,\ \theta = 0,\ \Delta\theta,\ 2\Delta\theta, ...,$ where $\Delta\varrho$ and $\Delta\theta$ are the sampling steps used for the quantization of the Hough space. For each $\mathbf{l}_k \in \mathcal{P}$ the values (ϱ_k, θ_k) of (5.21) are calculated and stored in the accumulator $\mathcal{A}(\varrho, \theta)$. In particular, for each value (ϱ_k, θ_k), the corresponding bin of the accumulator is incremented

$$\mathcal{A}(\varrho_k, \theta_k) = \mathcal{A}(\varrho_k, \theta_k) + 1.$$

The N_r accumulator bins with the highest count represent the parameters (ϱ_j, θ_j) of the N_r lines $\overline{\mathbf{x}}_{S_j}$, $j = 1, ..., N_r$, on the plenacoustic image. From (5.21) we have the lines

$$\begin{aligned} -\varrho_j + \cos(\theta_j) l_2 + \sin(\theta_j) l_3 &= 0, \\ -\varrho_j l_1 + \cos(\theta_j) l_2 + \sin(\theta_j) l_3 &= 0, \\ \tfrac{-\varrho_j}{\sin(\theta_j)} l_1 + \tfrac{\cos(\theta_j)}{\sin(\theta_j)} l_2 + l_3 &= 0, \\ \mathbf{x}_{S_j}^T \mathbf{l} &= 0, \end{aligned}$$

and the estimates of the source positions are, therefore,

$$\hat{\boldsymbol{x}}_{S_j} = \left[\frac{-\varrho_j}{\sin(\theta_j)}, \frac{\cos(\theta_j)}{\sin(\theta_j)} \right]^T,\ j = 1, ..., N_r. \tag{5.22}$$

As we can see, source localization and, in particular, the problem of disambiguating measurements and matching them with sources is here turned into a pattern analysis problem performed on an image. The fact that the patterns are linear, turns the localization algorithm into a linear procedure, which is quite a desirable feature.

5.9.2. Identification of edge rays

As in the case of a single reflector, the visibility of image sources is limited by the dimensions of the corresponding reflector. We determine the set of microphones that sense the reflective path from the image source $\hat{\boldsymbol{x}}_{S'_j}$, $j = 1, ..., N_r$ (N_r being the number of image sources) by testing if the microphone $\boldsymbol{x}_i$ receives energy from the direction of a given image source, i.e. we find

$$\mathcal{V}_{S'_j} = \{i \mid P_i(\mathbf{l}_j) > \beta \max(\mathcal{I}(\mathbf{l})),\ \hat{\mathbf{x}}_{S'_j}^T \mathbf{l}_j = 0\}.$$

Given the visibility information $\mathcal{V}_{S'_j}$ and the estimated image source position $\hat{\boldsymbol{x}}_{S'_j}$ the edge ray(s) are estimated as described in equations (5.8) and (5.9).

5.9.3. Estimation of visibility region edges

Given the set of edge rays $\mathcal{E}$ estimated from a number of plenacoustic images we want to estimate the visible endpoints of all reflectors in the environment. As the endpoints represent lines in the plenacoustic images, the edge rays cluster along lines in the reduced ray space as depicted in Figure 5.5. Each line represents the edge of the corresponding visibility region, i.e. the endpoint of the corresponding reflector in the geometric space. Therefore, we adopt, once again, the Hough transform in order to cluster the estimated edge rays and locate lines on the image. The procedure for the estimation of endpoint positions from the set of edge rays $\mathcal{E}$ is completely equivalent to the procedure described for the estimation of source positions (5.22) from the set of rays $\mathcal{P}$.

5.10. Considerations

In this Chapter we considered only specular reflections in order to estimate the reflector position and extension. However this does not mean that the estimation could not be performed in the presence of diffusion and diffraction. In presence of diffusive paths the microphone array receives rays originated from all points on the reflector. As a consequence, the presence of diffusive paths should "brighten up" the whole reflector's visibility region and reflector position and extension could be estimated from a single plenacoustic image without the need of a moving source. However, as already observed, the ideally diffusive surfaces are rare in acoustics

and, due to the limited resolution of the real plenacoustic camera, diffusive paths are hardly observable in presence of the much stronger specular reflections.

On the other hand, if present, diffraction could be used to extract the reflector's edge point directly from a single plenacoustic image. In fact, due to diffraction the edge acts as a secondary source, in the ray space the line corresponding to its position should "brightens up" and, therefore, the position of the diffraction edge could be estimated using the source localization procedure. However, in this dissertation no experiments were conducted to verify if the diffraction paths could be distinguished from the specular ones on the plenacoustic image, and it remains a topic for further research.

5.11. Results

In this section we show the results of some simulations and experiments to validate the applicability of the proposed approach and its performance.

As the geometric information does not depend on frequency, we use the wideband Minimum Variance Distortionless Response (MVDR) beamformer [75] to obtain the plenacoustic images (see also the Appendix C). In particular, the pseudospectra acquired in different frequency bands are used to compute the wideband pseudospectrum

$$P_i(\theta) = \prod_{k=1}^{K} P_{i,k}(\theta),$$

where k is the index of the sub-band centered at frequency f_k and $P_{i,k}(\theta)$ is the corresponding pseudospectrum. Furthermore, given the results of the resolution analysis performed in the section 4.4.1.2, in all simulations and experiments we subdivide the microphone array into sub-arrays with 3 microphones (i.e. $W = 3$).

In some of the tests plenacoustic images are shown. As described in Chapter 4 they are composed by a finite number of lines in the ray space. For clarity of visualization, however, images are shown over all the visible region of the plenacoustic camera by means of an interpolation procedure. Nonetheless, all the estimations are accomplished on non interpolated plenacoustic images. All simulations and experiments (except simulation 3 for reflector localization) use the LS estimator outlined in paragraph 5.6.

5.11.1. Multiple source localization

In order to numerically assess the accuracy of the above source localization algorithm we performed two simulations and three experiments:

- **Simulation 1** examines the performance of the localization of two sources for different values of the angular separation;
- **Simulation 2** examines the performance of the localization of two aligned sources;
- **Experiment 1** presents experimental results for localization of aligned sources;
- **Experiment 2** presents experimental results for localization of two sources at different distances;
- **Experiment 3** shows the experimental result of the challenging scenario with four acoustic sources.

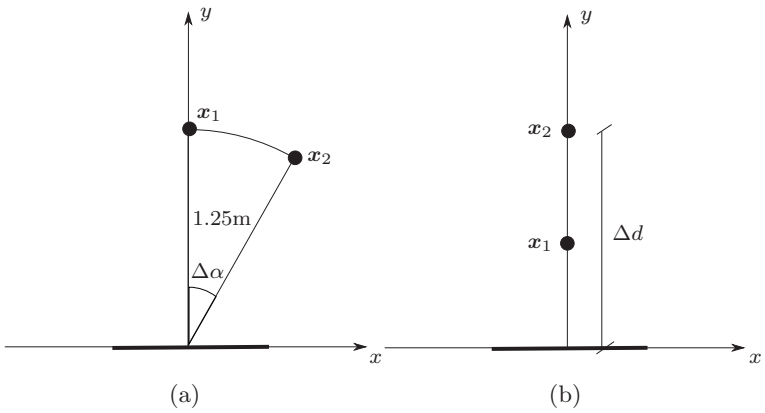


Figure 5.6.: Setup for the simulations to assess the accuracy of multiple source localization: (a) setup of the first simulation; (b) setup of the second simulation.

Simulation 1

The setup of the first simulation is shown in Figure 5.6 (a). The two sources, the first in $\boldsymbol{x}_1 = [0, 1.25\text{ m}]^T$ and the second in $\boldsymbol{x}_2 = [1.25\sin(\Delta\alpha), 1.25\cos(\Delta\alpha)]^T$, produce independent noises in the vocal bandwidth ($300 \div 4000$Hz). The signal acquired by the sensors is affected by an additive gaussian error with a SNR of

20 dB. Simulations have been performed with the ULA camera of 15 microphones spaced by 0.11 m.

An example of the acquired plenacoustic image is shown in Figure 5.7. The dashed lines represent the duals $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$ of the sources. Circles mark the peaks of the pseudospectra corresponding to the two sources. Crosses are located in correspondence of secondary peaks. In order to localize the multiple sources we need to distinguish between primary and secondary peaks and, at the same time, assign peaks to the corresponding sources. This can be readily accomplished using a Hough transform as described in section 5.9.1. The Hough transform, in fact, detects collinear local maxima and finds the parameters of the related lines, which become an estimate of the sources. In order to achieve sufficient accuracy in source localization, however, we need the grid density of the Hough map to be prohibitively large. This is why the Hough transform is here used only to find a first approximation of the source locations, which allow us to assign the peaks to the corresponding sources. A better estimate of the source locations can then be obtained through linear regression over measurements of the same source as described in section 5.4.

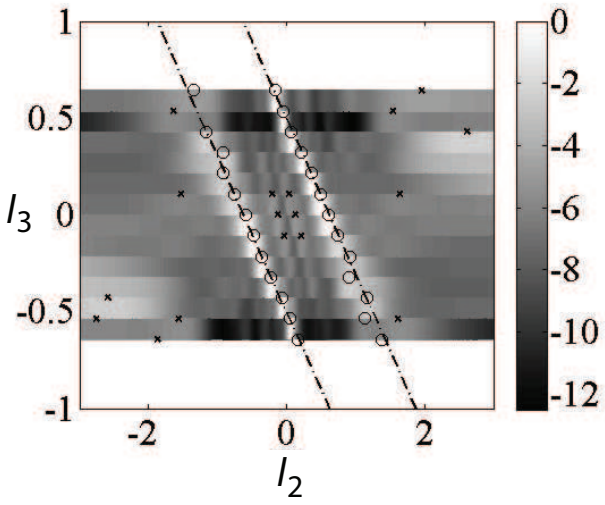


Figure 5.7.: Plenacoustic image with two sources.

The localization experiment is repeated for each location of $\boldsymbol{x}_2$ forty times, each with a different noise realization. Figure 5.8(a) and (b) show the localization error of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively, as $\Delta\alpha$ varies from 10° to 80°. The error on $\boldsymbol{x}_1$ is nearly constant. On the other hand, the localization of $\boldsymbol{x}_2$ is possible only for $\Delta\alpha \leq 60°$.

Beyond that angle the resolution loss becomes too relevant to guarantee a correct localization of peaks in $\mathcal{I}(\mathbf{l})$.

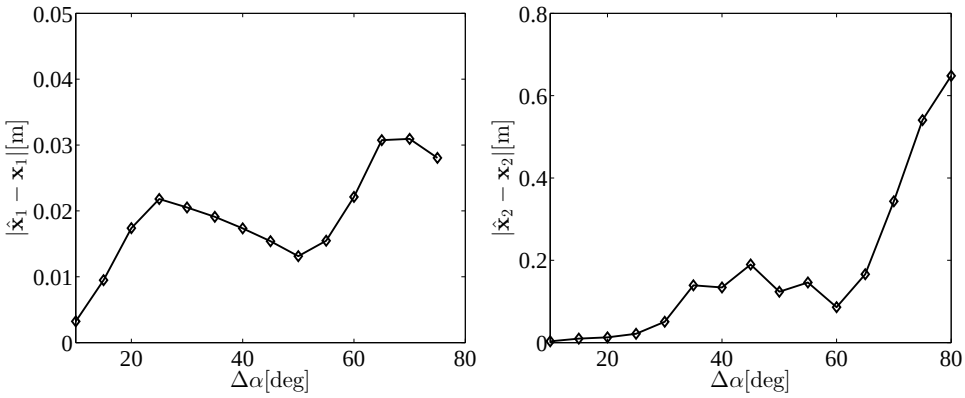


Figure 5.8.: (a) Localization error of $\boldsymbol{x}_1$ as a function of the angular difference $\Delta\alpha$ depicted in Fig.5.6(a); (b) Localization error of $\boldsymbol{x}_2$ as a function of the angular difference $\Delta\alpha$ depicted in Fig.5.6(a).

Simulation 2

Figure 5.6 (b) shows the setup of the second simulation. The sources are aligned on the line $x = 0$. The first source is in $\boldsymbol{x}_1 = [0, 1.25\,\text{m}]^T$, and the second one is in $\boldsymbol{x}_2 = [0, \Delta d]^T$. Δd ranges between 0.2m and 2.2 m. We use the same array configurations and source signals adopted for the previous simulation. Notice that in this setup the ray space representations $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$ of the two sources meet in $(l_1, l_2, l_3) = (1, 0, 0)$ (i.e. the ray corresponding to y axis), which makes the localization more challenging. An example of a plenacoustic image for sources aligned along the y axis is shown in the Figure 5.9. Localization techniques based on the traditional concept of acoustic camera would fail in this scenario due to mutual occlusion between sources. We show that plenacoustic imaging makes localization still possible for a wide range of distances Δd. Localization results are shown in Figure 5.10. The method guarantees a good localization accuracy when $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are not too close to each other. The error on $\boldsymbol{x}_2$ increases as it moves far away from the array, as the limited dimension of the observation window, compared to Δd, reduces the localization performance.

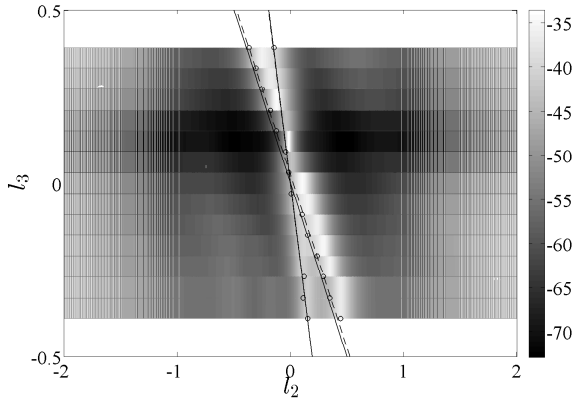


Figure 5.9.: For the sources aligned along the y axis the ray space representations $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_2$ of the two sources intersect in $(l_1, l_2, l_3) = (1, 0, 0)$.

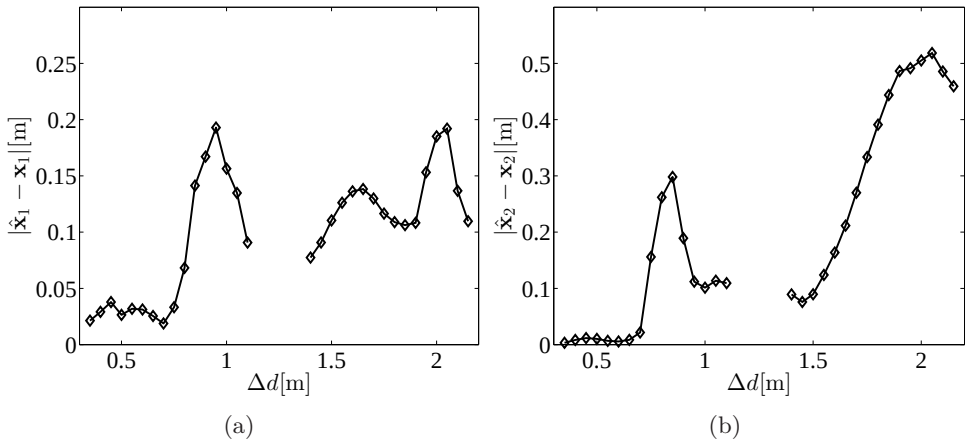


Figure 5.10.: (a) Localization error of $\boldsymbol{x}_1$ as a function of the distance Δd depicted in Fig.5.6(b); (b) Localization error of $\boldsymbol{x}_2$ as a function of the distance Δd depicted in Fig.5.6(b).

Experiment 1

We also conducted experiments to verify the accuracy of the algorithm on real-world data. All the experiments are conducted in a low-reverberation room with a ULA of 16 microphones spaced by 0.06 m. The first experiment follows the setup of Figure 5.6 (b), except for the fact that the first source is in $\boldsymbol{x}_1 = [0, 1\text{ m}]^T$, and the second one is in $\boldsymbol{x}_2 = [0, \Delta d]^T$, while Δd ranges between 1.4 m and 2.6 m. Localization results are shown in Figure 5.11. As seen in the above simulations, also in this case the localization improves as the distance between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ increases.

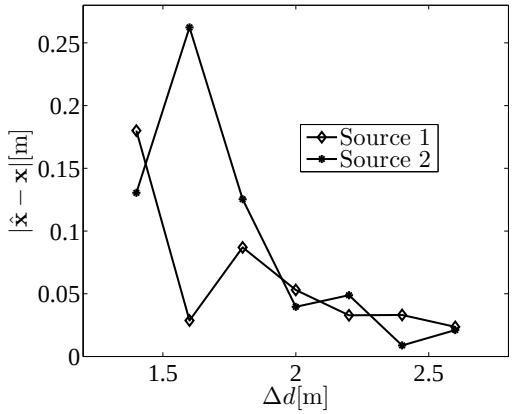


Figure 5.11.: Experimental results for the setup shown in Fig. 5.6(b).

Experiment 2

Figure 5.12 (a) shows the setup of the second experiment. Two acoustic sources are placed in $\boldsymbol{x}_1 = [\Delta x/2, 1.5\text{ m}]^T$ and $\boldsymbol{x}_1 = [-\Delta x/2, 1.5\text{ m}]^T$, respectively. The distance Δx between sources ranges from 0.2 m to 1.8 m. Results in Figure 5.12 (b) show that an accurate estimate is obtained even for $\Delta x = 0.2$ m, i.e. when the sources are very close to each other. As Δx increases the estimation error first diminishes and then increases again due to resolution loss.

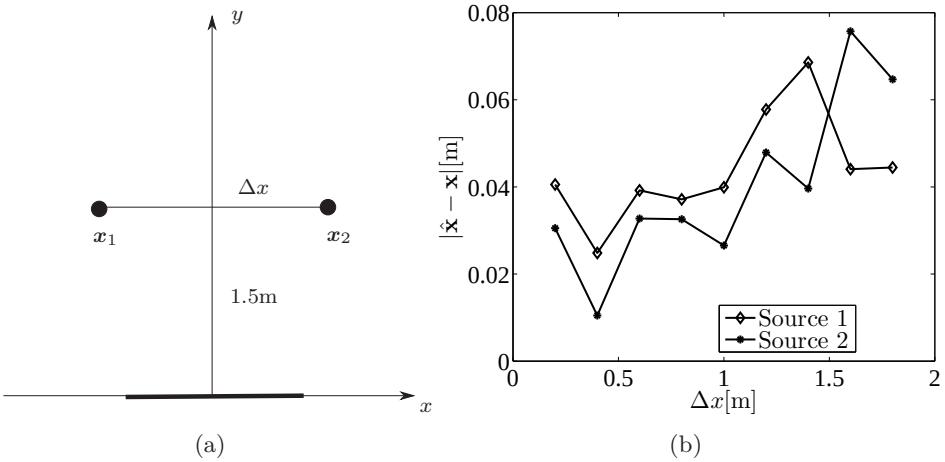


Figure 5.12.: (a) Experimental setup: two sources lie on a line that is parallel to the x axis at a varying distance Δx; (b) Results of the second experiment.

Experiment 3

In the third experiment we tested the system in a more challenging scenario of four acoustic sources. The setup and the estimated source positions are shown in Figure 5.13 (a)(b). Crosses and circles mark the estimated and actual locations of the sources, respectively. Figure 5.13 (d) shows the acquired plenacoustic image. In order to assess how well real data match simulative data, Figure 5.13(c) shows the simulated plenacoustic image. Notice that the modeled and acquired plenacoustic images are, in fact, very similar, thus confirming the validity of the model. The algorithm is able to correctly discriminate between contributions of different sources and estimate their positions due to the fact that the corresponding peaks naturally cluster on lines on the plenacoustic image, as shown in Figure 5.13(d). As expected, the estimation accuracy is better when the sources lie closer to the center of the array.

5.11.2. Reflector localization

Tests are organized in seven different sessions:

- **Simulation 1** shows the performance of reflector line estimation as a function of the distance between camera and reflector;
- **Simulation 2** is aimed at verifying the accuracy of the reflector endpoint estimator as a function of the number of microphones;
- **Simulation 3** compares analytical and simulated performances of the algorithm at different values of the signal-to-noise ratio (SNR) and the gain obtained using the iterative algorithm described in paragraph 5.8;
- **Simulation 4** compares the performance of the proposed algorithm with a state-of-the-art technique;
- **Simulation 5** is an example of localization of multiple reflectors;
- **Experiment 1** consists in reflector line and reflector endpoint estimation in a real world scenario;
- **Experiment 2** shows the applicability of the method in more complex environments.

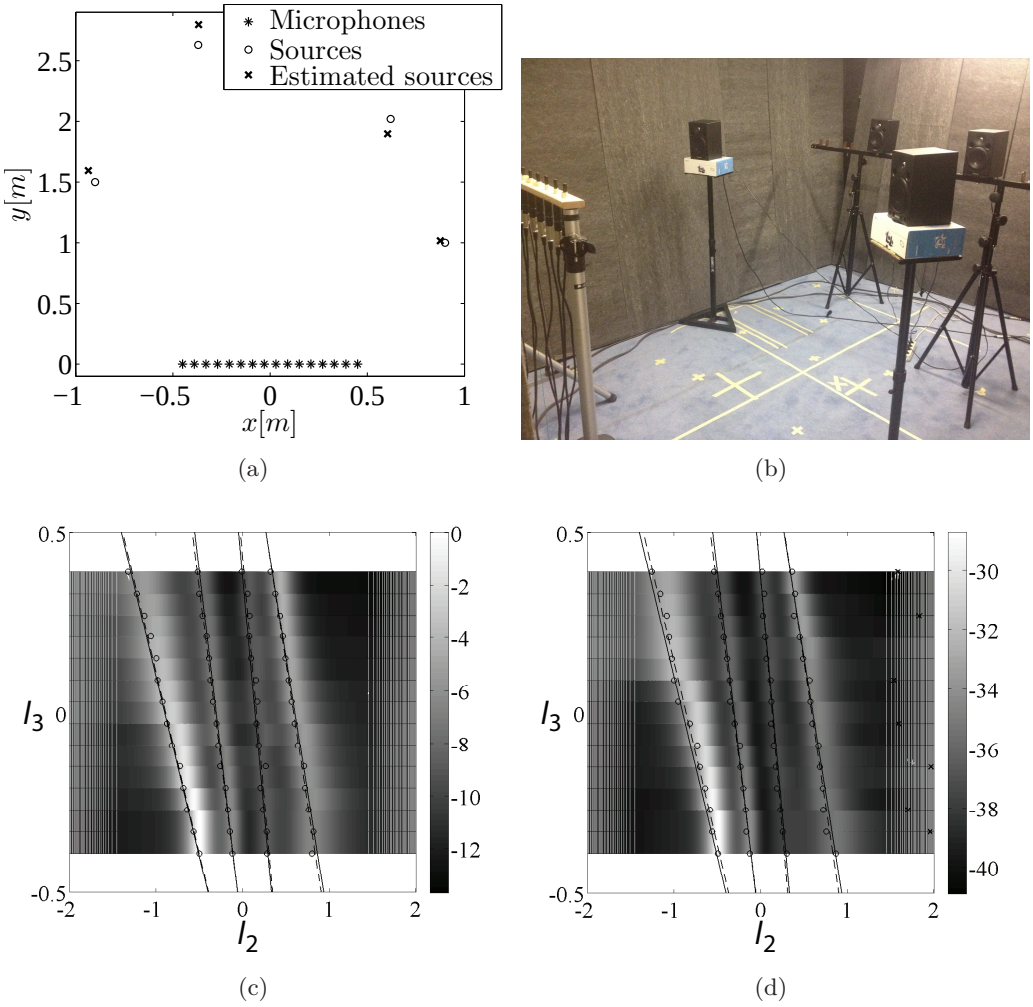


Figure 5.13.: (a)(b) Experimental setup with four sources present in the acoustic scene; (c) Modeled plenacoustic image; (d) Acquired plenacoustic image.

Simulation 1

We tested the accuracy of the reflector localization algorithm through simulations based on the setup of Figure 5.14. The source is in $\boldsymbol{x}_S = [0, 0.5\text{m}]^T$, and the reflector is at a distance D from the x axis, which ranges from 0.6 m to 1.5 m, and it is parallel to it. The source produces a white noise with the bandwidth ($300 \div 4000$Hz) with a SNR of 20 dB. We conducted the test using the ULA camera of 15 microphones spaced by 0.11 m. For the computation of the plenacoustic image we use $W = 3$.

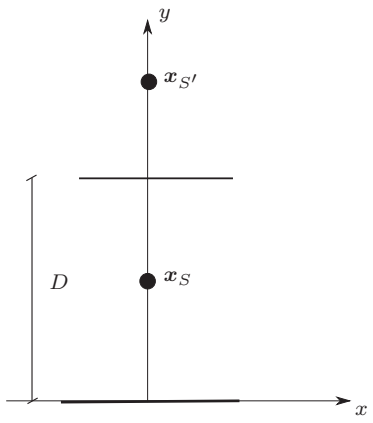


Figure 5.14.: Simulation 1 – setup.

Figures 5.15 (a) and 5.15 (b) plot the localization error of $\boldsymbol{x}_S$ and $\boldsymbol{x}_{S'}$, respectively. Figures 5.15 (c) and 5.15 (d) plot the error on the distance and the angle of the estimated reflector with respect to the actual one, respectively. The error on $\boldsymbol{x}_S$ is nearly constant for all the distances. A different situation arises for $\boldsymbol{x}_{S'}$. In fact, when D is below 0.7 m, $\boldsymbol{x}_S$ and $\boldsymbol{x}_{S'}$ are close each other, and the algorithm exhibits a poor accuracy on localizing $\boldsymbol{x}_{S'}$. For intermediate distances the localization error decreases. If D is above 1.2 m the error on $\boldsymbol{x}_{S'}$ becomes larger, due to the limited extension of the array with respect to D.

Simulation 2

In the second test we test the accuracy of the reflector endpoints localization as a function of the number of microphones in the array. Sensors are spaced by $d = 10$ cm. At disposal we have two speaker arrays, each one composed of 5 loudspeakers spaced by 20 cm, placed behind the microphone array at distance 2 m, as in Figure 5.16. The environment is probed with a white noise signal in the bandwidth [1 kHz − 10 kHz] with a SNR of 20 dB.

Using analytical expressions for LS estimate (5.17) ($\mathbf{C}_\theta$ is estimated from simulations) we obtain the expected variances in function of variable M. Results are shown in Figure 5.17 for both x and y coordinates. As expected, the accuracy increases with the number of sensors, showing also that the accuracy is higher for the x coordinate rather than for the y one. This is due to the fact that for the geometry adopted the x coordinate is related to the direction of arrival of the source,

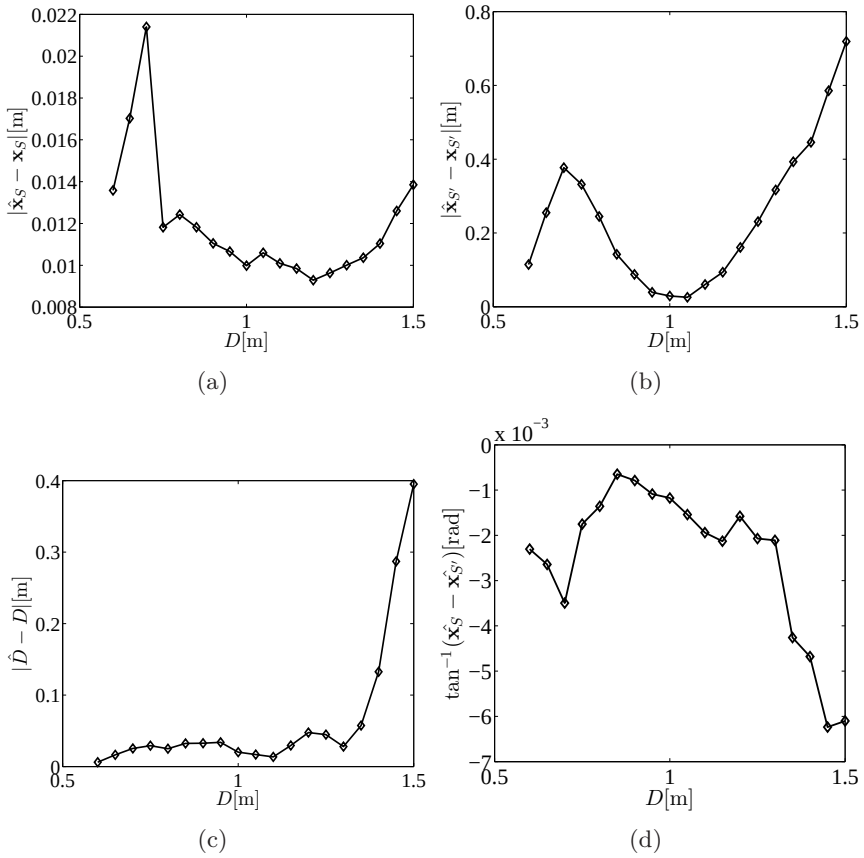


Figure 5.15.: Simulation 1 – localization error of image sources and of the reflector for the setup in Figure 5.14: (a) direct source $\boldsymbol{x}_S$; (b) image source $\boldsymbol{x}_{S'}$; (c) distance error; (d) angle error.

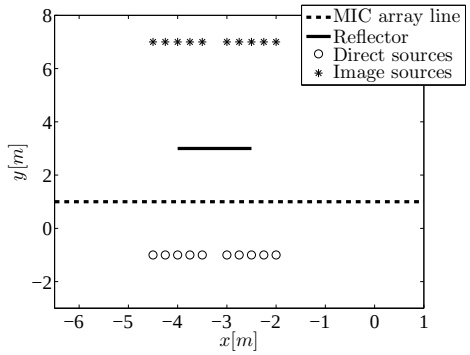


Figure 5.16.: Simulation 2 – system configuration

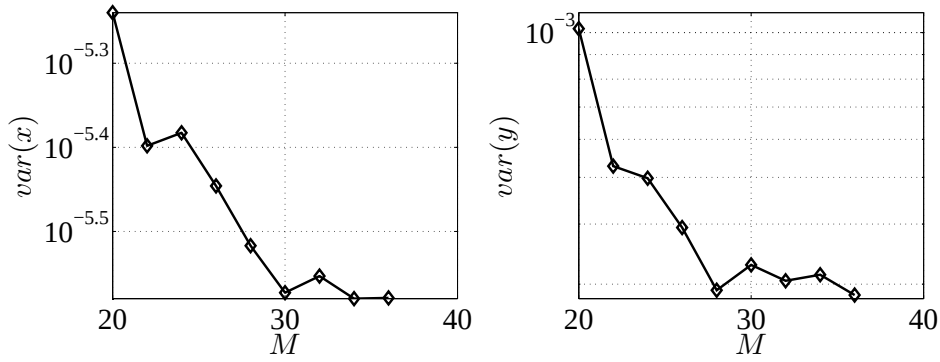


Figure 5.17.: Simulation 2 – endpoint estimation variance (log scale) in function of the number of microphones; being the system completely symmetric only the data for a single endpoint is visualized.

while the y coordinate coincides with its range. It is known that linear arrays better discriminate the DOA rather than the range.

We have also conducted 1000 independent simulations in order to validate the analytical results. In simulations microphones and loudspeakers are modelled with cardioid patterns directed towards the reflector. Data for 26 microphones are reported in the Table 5.1.

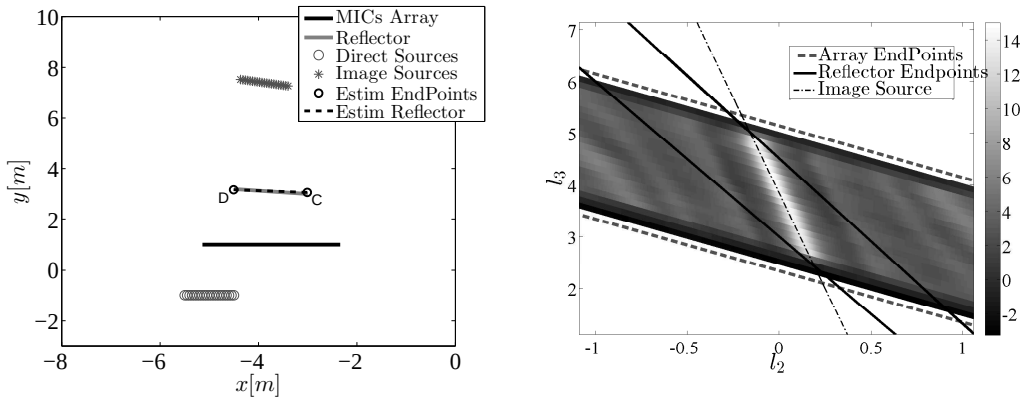
Table 5.1.: Simulation 2 – theoretical and simulated variance.

Variance	x [m]	y [m]
Theoretical (26 Mics)	$3.582 \cdot 10^{-6}$	$3.933 \cdot 10^{-4}$
Simulated (26 Mics)	$3.357 \cdot 10^{-6}$	$3.643 \cdot 10^{-4}$

Simulation 3

In the third test we analyze the reflector endpoint estimation for different values of SNR. Moreover, we compare the accuracy of the LS technique in paragraph 5.6 with that of the iterative WLS methodology outlined in paragraph 5.8. The configuration is shown in Figure 5.18 (a). The array is composed of 25 sensors. Figure 5.18(b) shows an example of the obtained plenacoustic images (in the reduced ray space with $l_1 = 1$) for a single source position and $SNR = 20$ dB. The reflector estimate is obtained moving the source at 21 different positions in space is shown in Figure 5.18 (a).

In Figure 5.19 the theoretical variance of endpoint estimates obtained using analytical expressions for the LS technique; the simulation variance estimated from



(a) (b)

Figure 5.18.: Simulation 3: (a) the configuration in the geometric space with estimation example (black dotted line); (b) example of the measured acoustic image shown in the reduced ray space.

1000 independent realizations for the LS technique; the simulation variance for one iteration of the WLS algorithm; and the theoretical variance for the ML technique are shown for different values of SNR. The figure sheds light on the following facts:

- the theoretical variance well approximates the variance obtained using simulations;
- the iterative WLS algorithm achieves a significant gain with respect to LS technique;
- one cycle of the iterative algorithm approaches the performance of the ML technique and as a consequence the gain obtained by successive iterations is reduced.

We recall, however, that the iterative algorithm, in order to achieve significant gain, requires the knowledge of $\mathbf{C}_\theta$ which is here estimated from the simulations.

Simulation 4

The comparison with other methods is not easy because of different assumptions and hardware used. Furthermore no other method, to our knowledge, enables the estimation of reflectors endpoints. A numerical comparison is therefore out of the scope.

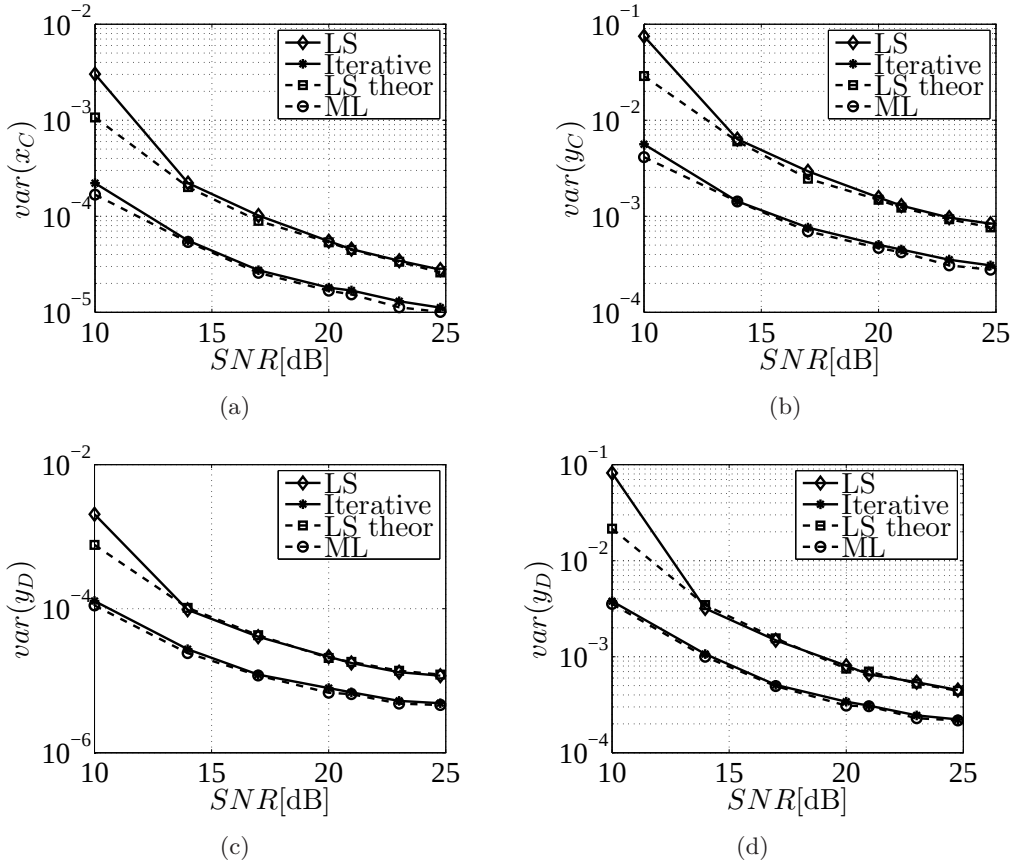
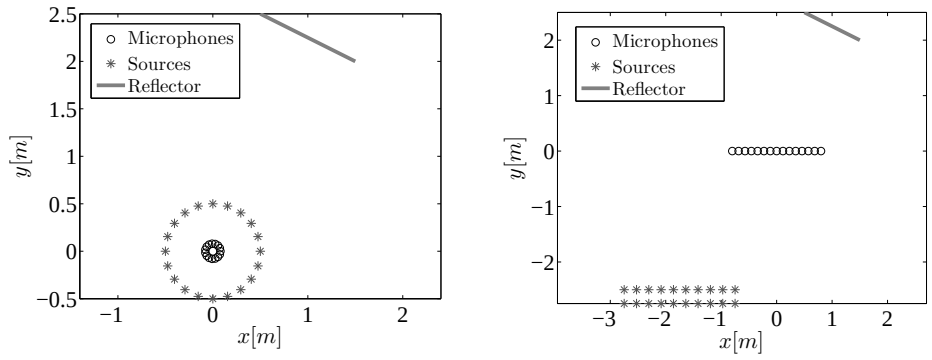


Figure 5.19.: Simulation 3 – the theoretical variance obtained using analytical expressions for the LS technique; the simulation variance estimated from 1000 independent realizations for the LS technique; the simulation variance for one cycle of the iterative WLS algorithm; and the theoretical variance for the ML technique.

Among all the techniques in the literature, the most similar is [5], in fact in both methods similar hardware (a microphone array and a moving acoustic source) and same information extracted from the acquisitions (both use beamforming techniques to identify DOAs) are used. However, in order to work, the technique in [5] requires a compact array. Therefore different configurations are used, which makes a direct comparison not possible. The simulation configurations are shown in Figure 5.20. In Figure 5.20 (a) a circular array with radius 0.08 m is used to extract constraints on reflectors position using multiple source locations. In Figure 5.20 (b) a 1.6 m long linear array is used in order to extract plenacoustic images for different source positions using the method discussed in this Chapter. The number of sensors is 15



(a) (b)

Figure 5.20.: Simulation 4: (a) a circular array and a moving source is used to estimate the position of the reflective surface; (b) a linear array and a moving source is used to estimate the position of the reflective surface.

for both arrays, sources emit a white noise in band 1 kHz − 10 kHz and SNR is 20 dB. The wideband MVDR method is used for computing pseudospectra in both cases.

We perform the estimation process using a different number $(5 - 20)$ of source positions randomly chosen among the 20 source positions shown in Figure 5.20 (a) and (b). As the method proposed in [5] assumes that the array reflective paths are present for all the sensors, the source positions in Figure 5.20 (a) are chosen so that the visibility is always guaranteed. On the other hand, the sources used in Figure 5.20 (b) occupy a wider area as there is no constraint on complete visibility. Furthermore, method based on DOA constraints assumes that the source position is known.

Figure 5.21 shows the estimation bias and variance for: the distance of the reflector from the origin of the reference frame (center of the array); the angle of view under which the reflector line is seen form the origin of the reference frame. Bias and variance are estimated from 100 independent simulations for all the possible values of source locations. Using multiple source positions we combine multiple independent observations and therefore, as expected, the estimation variance decreases linearly with the number of positions (e.g. angle variance is $16{\cdot}10^{-4}$ deg for 5 source positions and $4 \cdot 10^{-4}$ deg for 20 source positions). The method based on DOA constraints

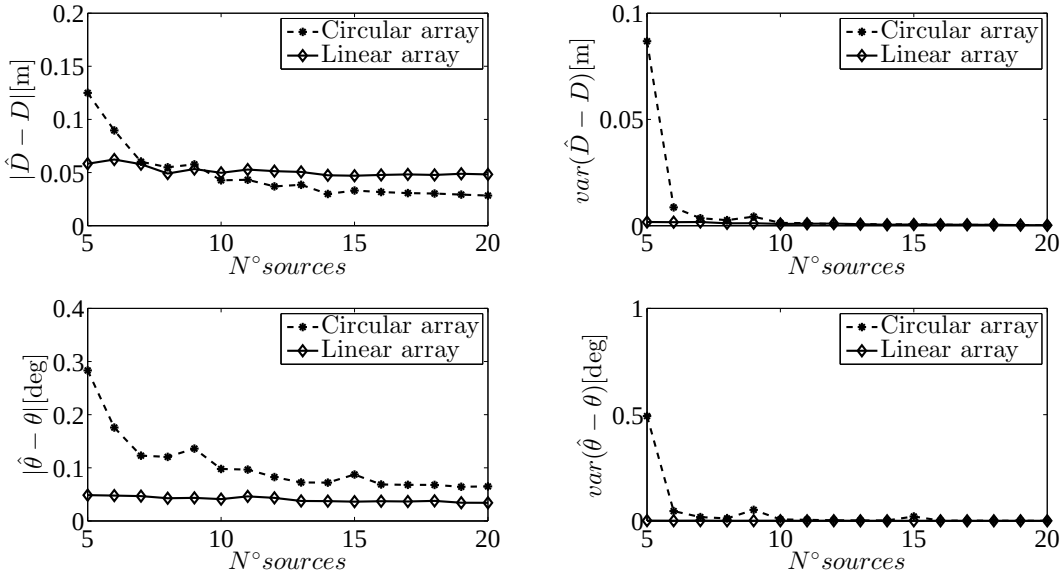


Figure 5.21.: Simulation 4 – estimated distance/angle bias and variance shown for circular and linear array obtained using 100 independent simulations.

converges rapidly and the performance of two methods is comparable for higher number of source positions.

Table 5.2.: Simulation 4 – comparisons of two estimation methods.

	DOA [5]	*Plenacoustic*
Distance (bias) [m]	0.0283	0.0482
Distance (variance) [m]	$1.7662 \cdot 10^{-4}$	$1.9411 \cdot 10^{-4}$
Angle (bias) [deg]	0.0650	0.0342
Angle (variance) [deg]	$5.9414 \cdot 10^{-4}$	$4.1490 \cdot 10^{-4}$
x_C (bias) [m]	–	0.0291
y_C (bias) [m]	–	0.0395
x_C (variance) [m]	–	$0.0368 \cdot 10^{-3}$
y_C (variance) [m]	–	$0.1609 \cdot 10^{-3}$
x_D (bias) [m]	–	0.0157
y_D (bias) [m]	–	0.0455
x_D (variance) [m]	–	$0.0252 \cdot 10^{-3}$
y_D (variance) [m]	–	$0.1684 \cdot 10^{-3}$

While the performance of the two approaches is generally similar notice that the method based on DOA constraints requires full visibility and the knowledge of the

source location, whereas the method presented in this Chapter can work with limited visibility and unknown source positions. Furthermore, the reflector endpoints can be estimated as well. Table 5.2 summarizes the results obtained using 20 source positions and 100 realizations.

Simulation 5

We now show some results obtained using a configuration with multiple reflectors. The simulation examples for cases with two and three reflectors are shown in Figures 5.22 (a)(b) and (c)(d) respectively. Figure 5.23 shows the Hough transforms for images in Figures 5.22 (b) and (d) used to cluster rays coming from the same image sources.

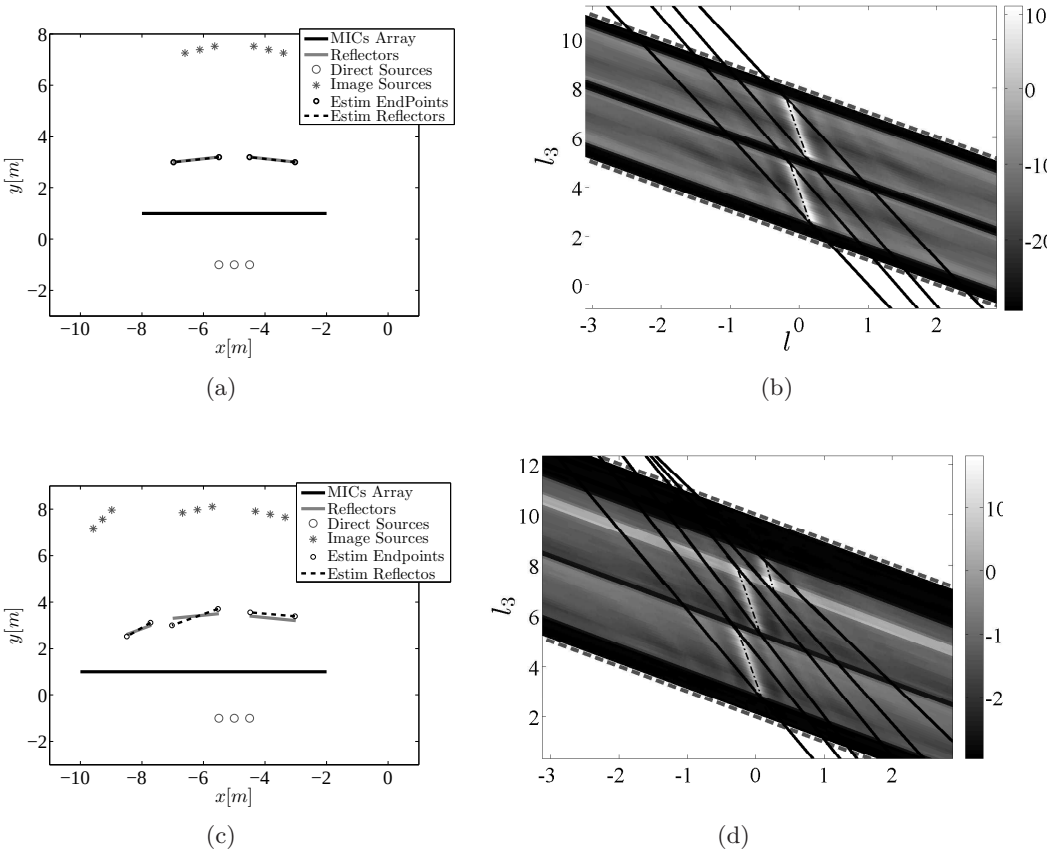


Figure 5.22.: Simulation 5: (a)(c) the configuration in the geometric space with estimation examples (black dotted lines); (b)(c) example of the measured acoustic image shown in the reduced ray space; for environments with two (a)(b) and three (c)(d) reflective surfaces.

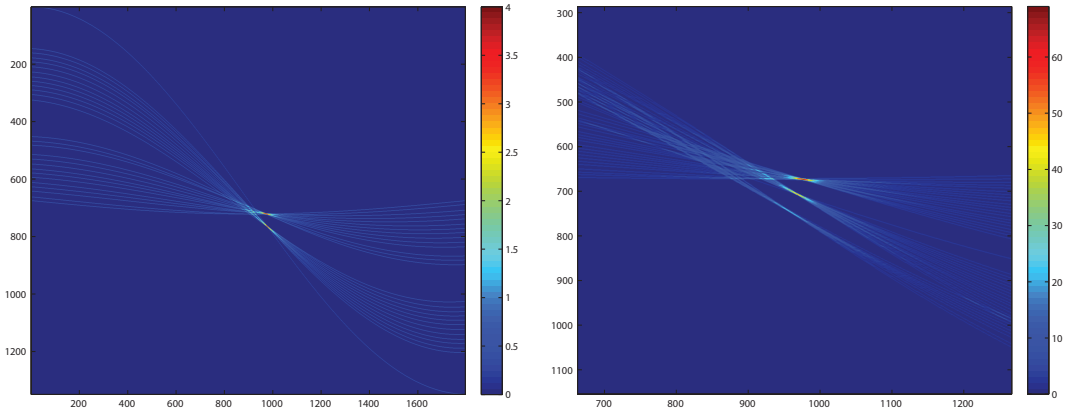


Figure 5.23.: Simulation 5: (a) Hough transform of the image in Figure 5.20 (b); (b) Hough transform of the image in Figure 5.20 (d).

Notice that when three reflectors are present the accuracy degrades significantly with respect to the case of two reflectors. The reason can be found in the fact that plenacoustic images become densely populated of image sources, which partially overlap, thus leading to a wrong localization.

(a) (b)

Figure 5.24.: Experimental setup for: (a) experiment 1; (b) experiment 2.

Experiment 1

In this experiment we use a logarithmically spaced linear array composed of 13 microphones. The experiment has been conducted in a low-reverberation room and

the signals acquired by the sensors were sampled at $F_s = 44.1$ kHz. The environment was "illuminated" by a small loudspeaker emitting a white noise in the frequency band 1 kHz $-$ 10 kHz. We used a $W = 3$ spatial rectangular window to perform beamforming.

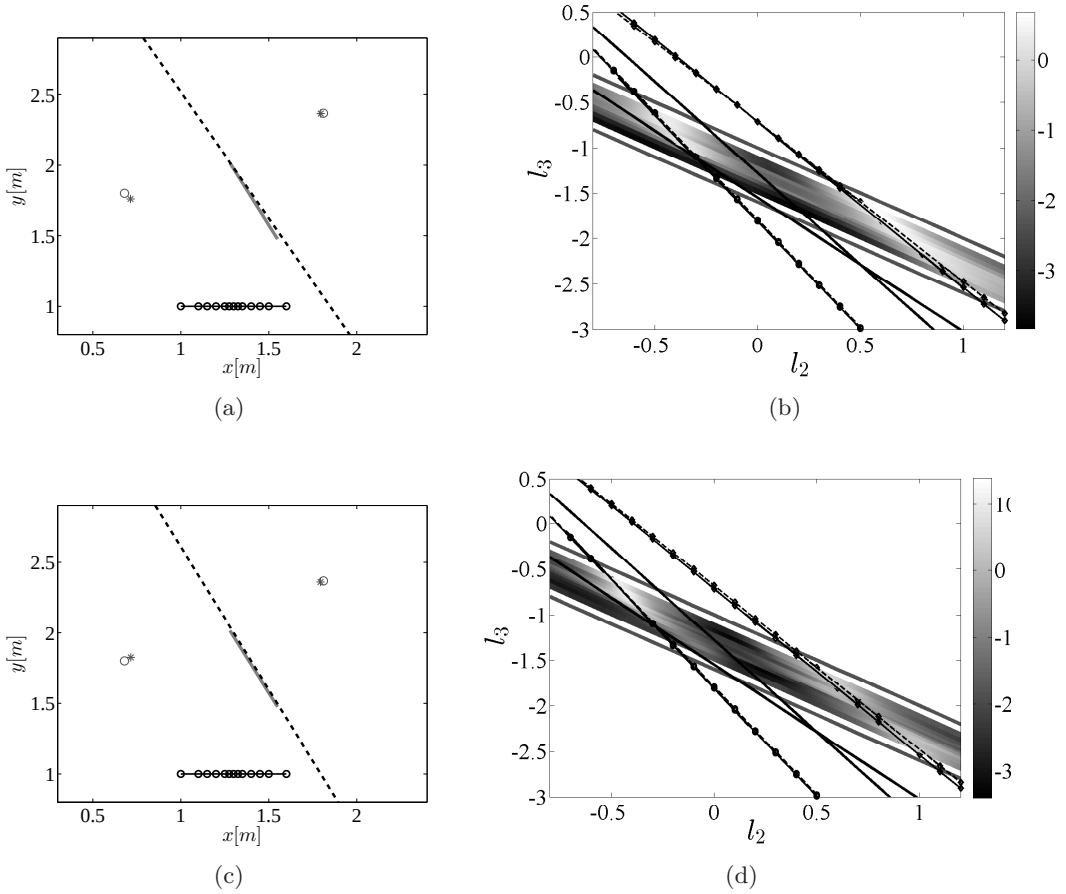


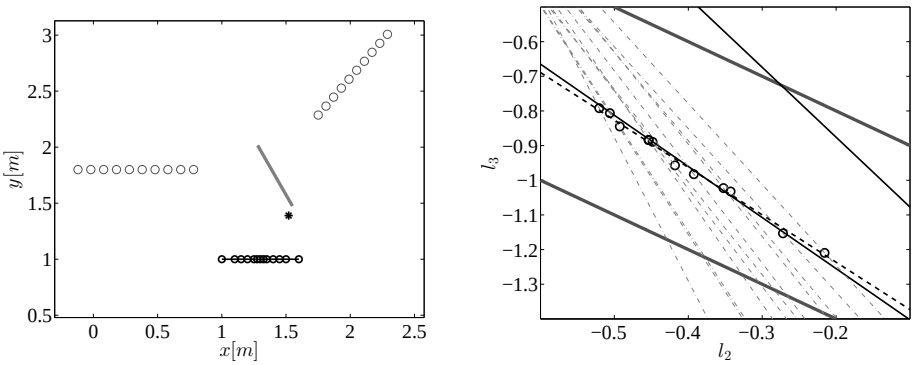
Figure 5.25.: Experiment 1: (a) configuration and simulations estimates in the geometric space; (b) configuration and simulations estimates in the ray space; (c) configuration and experiment estimates in the geometric space;(d) configuration and experiment estimates in the ray space; circles and stars are the direct/image sources and their estimates, respectively; continuous and dashed lines are the reflector and the estimated lying line of the reflector, respectively.

A 60 cm long reflective surface is placed in front of the array as shown in Figure 5.24 (a). The presence of the reflector causes reflections to appear. The array is then used to estimate the position of both real and image loudspeakers. This estimates can then be used to estimate the line on which the reflector lies. The results of the experiment are shown in Figure 5.25 (c) and (d) and are compared with simula-

tion results shown in Figure 5.25 (a) and (b). As expected, the aperture problem and the relatively small total number of microphones causes the plenacoustic image to be blurred (aperture). Yet, the image clearly exhibits the lines that represent the loudspeaker and its (wall-reflected) image. In addition, the plenacoustic image clearly shows the visibility of the image loudspeaker (which is to be confined within the boundaries of the reflector). The estimation results are summarized in the Table 5.3. Even in the presence of noise; measurement errors on reference positions; finite aperture of the array; and discrete sampling of the ray space, Table 5.3 shows a good estimation accuracy and a good match between simulations and real world experiments.

Table 5.3.: Experiment 1 – source and reflector estimates.

	Source $[x, y]$ $[m]$	*ImageS* $[x, y]$ $[m]$	*Reflector* $[l_1, l_2, l_3]$
Real	$[0.68, 1.8]$	$[1.811, 2.366]$	$[0.3925, 0.1966, -0.8985]$
Sim.	$[0.7141, 1.759]$	$[1.796, 2.364]$	$[0.3751, 0.2097, -0.9030]$
Exp.	$[0.7159, 1.825]$	$[1.794, 2.359]$	$[0.3925, 0.1943, -0.8990]$



(a) (b)

Figure 5.26.: Experiment 1: (a) Geometric space – source positions, marked with gray circles, and the estimated endpoint, marked with black star; (b) Ray space – source positions, represented as gray dotted lines, and corresponding edge ray estimates, marked with black circles, used for endpoint estimation; endpoints are represented with black lines and their estimates with black dotted lines.

Finally the reflector endpoints can be estimated if a number of edge rays are observed (theoretically ≥ 1 if the reflector line estimate is used as an edge ray and ≥ 2 if not as in this case). This can be achieved moving the acoustic source. The example for the estimation of a reflector endpoint obtained using images of 10 different positions is shown in Figure 5.26. Figure 5.26 (b) shows source position

estimates and corresponding edge ray estimates used for endpoint estimation.

Experiment 2

The last experiment was performed using using a 0.9 m long uniform linear array composed by 16 microphones as shown in Figure 5.24 (b). The geometry of the setup, along with the estimated geometry, are shown in Figure 5.27 and results are summarized in the Table 5.4.

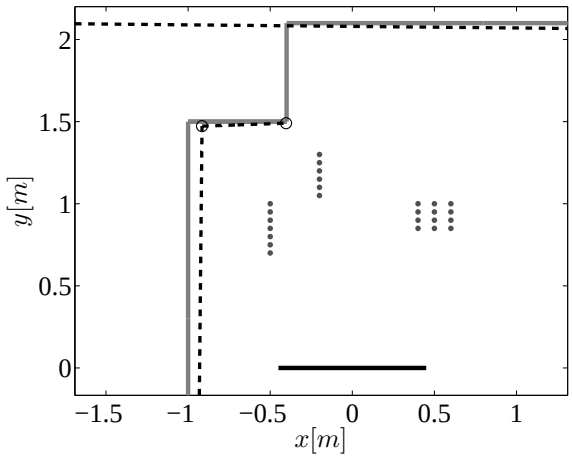


Figure 5.27.: Experiment 2: source positions (dots); reflectors (continuous lines); reflector line estimates (dotted lines) and reflector endpoint estimates (circles).

Table 5.4.: Experiment 2 – endpoint and reflector estimates.

	Real	*Estimate*
Point [m]	$[-0.4, 1.5]$	$[-0.4037, 1.491]$
Reflector 1	$[0.7071, 0, 0.7071]$	$[0.7323, -0.0073, 0.6810]$
Reflector 2	$[0, 0.5547, -0.8321]$	$[-0.0196, 0.5533, -0.8327]$
Reflector 3	$[0, 0.4299, -0.9029]$	$[0.0041, 0.4333, -0.9012]$

Notice that the estimation of a fourth reflector was not possible as it does not generate first-order reflections visible from the microphone array. It could be estimated using second order reflections but, due to resolution limits, higher order reflections are not considered in this work.

5.12. Conclusive remarks

In this Chapter we presented a novel method for geometry inference based on the plenacoustic images introduced in the Chapter 4. The proposed framework made use of a long (or moving) microphone array but it can be easily applied to a spatial distribution of microphone arrays with a consequent gain in accuracy as the "positions of observation" of the acoustic scene increase. Although plenacoustic image analysis may seem quite demanding compared to other geometric inference algorithms, the advantages in using this method for environment estimation are quite relevant. The nonlinear transformation that maps acoustic measurements onto the ray space renders the estimation constraints linear, which means that there is no need of iterative minimization procedures, and the solutions can be found using a least-squares method. Furthermore, the high regularity of the ray space allows an easy distinction of different acoustic events. So the multiple acoustic sources are separated and localized by simply finding lines on the plenacoustic image. Furthermore, through this method it is generally possible to detect reflectors of smaller size, and in positions that would make it hard for other methods to succeed. Visibility discontinuities do not represent a problem and, when observed, they can actually help estimate not just the line where the reflector lies upon, but also the spatial extension of the reflector. In particular, the extension of the reflector can be estimated if the endpoints of the reflector are visible from the spatial window occupied by the microphone array. This property, in fact, becomes useful in estimating irregular environments that are characterized by occlusions and limited visibility of acoustic reflections.

Chapter 6

Application to environment awareness: reflection coefficient estimation

THIS Chapter presents a novel model-based method for in-situ estimation of sound reflection coefficients in acoustic enclosures. The method uses the modeling engine presented in the Chapter 3 for simulation of acoustic propagation and generation of the spatial pseudospectrum using a beamforming technique. The simulated pseudospectrum is matched with the measured pseudospectrum acquired by a real microphone array in order to estimate reflection coefficients of the walls of the acoustic enclosure. In particular, the reflection coefficients and the source signal amplitude are estimated by finding optimum model parameters, such that both the simulated and real pseudospectra are closely matched. Experimental results confirm that the proposed method allows for an accurate estimation of the reflection coefficients, especially for reflective walls, typical of everyday environments.

In this Chapter we use a compact circular array, which is a traditional acoustic camera, as rays are bound to pass through the center of the array. Nonetheless, the information captured by this array, i.e. the spatial pseudospectrum (called also angular pseudospectrum), is the same used to acquire plenacoustic images in Chapter 4. Therefore, although acquired with an extremely narrow (plen)acoustic cam-

era, the use of the pseudospectra for reflection coefficient estimation shows effectively that the plenacoustic images capture not only the geometric information (extracted in Chapter 5) but also the radiometric information about the environment.

6.1. Motivations

As already stated, many audio signal processing algorithms, such as signal enhancement [99] and multichannel upmixing [100], can benefit from the knowledge of parameters characterizing the acoustic environment. Recent research efforts focused primarily on estimating the Directions Of Arrival (DOAs) of early room reflections [101, 81, 102], room volume estimation [103], and the inference of the geometry of an acoustic enclosure [35]. In Chapter 5 we presented a novel approach for geometry inference. In this Chapter we focus on the problem of estimation of reflective properties of walls in an environment. The knowledge of the acoustic properties of walls in an environment is crucial for many applications in the realm of space-time audio processing. Consider, for example, the problem of rendering a virtual acoustic source by means of a loudspeaker array. When the system is operating in a reverberant environment reflections from obstacles could severely corrupt the spatial impression, if they are not properly accounted for. If, on the other hand, the location of walls and their reflective properties are known, reverberation can be cancelled out up to a certain degree [83]. Early reflections can even be exploited to enhance the spatial impression [53], [104].

When a sound wave propagating through an acoustic space is reflected from a wall, a part of the impinging energy is reflected and another part is absorbed. Such reflective properties of room boundaries are captured by reflection coefficients, which are typically measured using a boundary sample positioned in the so-called impedance tube [105]. However, direct estimation of the reflection coefficient values in a room is highly challenging due to the high sensitivity to noise and the dependency of the reflection coefficient on both frequency and the angle of incidence.

6.2. Related work

In this Chapter, we propose a methodology for a sub-band *in-situ* estimation of the reflection coefficient of planar obstacles. This is a challenging task, as the re-

flection coefficient depends on the frequency and angle of incidence and, moreover, the estimate is highly sensitive to background noise and the presence of interfering sources during the measurement procedure. For this reason, accurate reflection coefficient estimations are typically conducted in acoustically conditioned environments. In particular, several techniques that can be found in the literature are based on the use of impedance tubes [105]. A sample of the investigated material is placed at one end of the tube, and the other end is equipped with a loudspeaker. The loudspeaker emits a narrowband signal, which is tuned to produce standing waves in the tube. The ratio of the sound pressure at peaks and troughs is measured, which enables the computation of the reflection coefficient. This technique has been standardized by standard methods ISO 10534-1 (for narrowband estimation) and ISO 10534-2 (for wideband estimation). Despite high accuracy, the impedance tube measurements are costly and complex to be carried out, as a portion of material needs to be removed. Its application for multimedia installations is therefore not always possible.

Alternatively to the impedance tube methods, reflective properties can be measured in a reverberant chamber. The simplest method of this kind is based on the Sabine's equation [105], which relates the reverberation time of the environment, its volume and the average absorption coefficient of the walls. The estimation of the reverberation time when the material under test is present and when it is absent in the reverberation chamber, enables the estimation of its reflection coefficient. It is worth noticing that this technique requires that the whole measurement procedure is accomplished in controlled conditions, and therefore it can be costly for many applications.

A different route is pursued for *in-situ* measurements. Relevant examples are [106] and [107]. In [106] the author measures the Acoustic Impulse Response (AIR) of the environment. The AIR consists of the direct path between the source and the microphone, the echo related to the reflection coming from the material under test and all other reflections. Under the assumption of being temporally distinct from the echo related to the material under test, the direct signal and reverberation are canceled out from the AIR through a temporal windowing. The reflection coefficient is then estimated by comparing the amplitude of direct and reflected echoes. This measurement procedure has been standardized in ISO 13472-1. Although this technique is attractive due to simplicity, it can fail when the echo related to the material under test overlaps with other echoes (e.g. think about a measuring sys-

tem placed in the center of a square room). Furthermore, an efficient cancelation of the direct echo requires an accurate modeling of the measurement device. In [107] the measurement microphone is replaced by a microphone array, placed between the probing loudspeaker and the material under test. The spatial response of the array is parameterized so that the maximum directivity can be focused towards the loudspeaker (and minimum towards the material under test) or conversely, the spatial response can exhibit a maximum towards the material and a minimum towards the sound source. This way the array allows to separate the direct path signal from the reflected one, while limiting the influence of reverberation. The reflection coefficient is then estimated by comparing the amplitudes of direct and reflection paths, similarly to [106]. When the environment is composed of walls of different materials, both [106] and [107] require multiple measurements, one for each wall, which makes these procedures time consuming.

The technique proposed in this Chapter, on the other hand, aims at estimating, in a sub-band fashion, the reflection coefficients of the walls through a single measurement. The proposed method makes use of a microphone array and of an acoustic source signal. A beamforming technique allows to estimate the angular distribution of the energy incoming at the microphone array. Under the assumption that the geometry of the environment, the position of the source, its spatial response, the position of the microphone array and its spatial response are known, the angular pseudospectrum can also be modeled. In particular, it results in a linear combination of individual contributions, which are known up to a scale factor, which depends non-linearly on the reflection coefficients. The estimation is performed in two steps. The first aims at estimating the individual scaling factors of the contributions of the pseudospectrum, whereas the latter infers the reflection coefficients from such scaling factors. The proposed technique turns out to be robust against noise and interference, thus paving the way for applications in multimedia and all the scenarios where cost is an issue.

6.3. Problem formulation

Reflection coefficient of the material is defined as the amplitude of a reflected wave relative to an incident wave [105]. In general, the reflection coefficient value $\beta_k(f, \theta_k)$ for reflector k depends on frequency f and angle of incidence θ_k. We assume here that all surfaces are planar and incident waves are reflected from walls in a specular

fashion, i.e., the angle of incidence is equal to the reflection angle. As illustrated in Figure 6.1, upon specular wave reflection, the reflected signal amplitude is attenuated by the value of the travelled distance d and the reflection coefficient value $\beta_k(f, \theta_k)$.

Let us now separate the reflection coefficient into two components charaterizing its dependency on f and θ_k by rewriting

$$\beta_k(f, \theta_k) = \alpha_k(f) g_k(\theta_k), \tag{6.1}$$

where $\alpha_k(f)$ denotes the reflection coefficient of the reflector k at frequency f and $g_k(\theta_k)$ models the variation of the reflection coefficient with the angle θ_k. In particular we are interested at estimating $\alpha_k(f)$, whereas $g_k(\theta_k)$ is assumed to be known. Even if $g_k(\theta_k)$ is unknown we will suppose it constant and estimate only the frequency dependent factor $\alpha_k(f)$.

Note that the amplitude of the signal acquired by a microphone depends also on the amplitude A of the emitted signal $s(t)$, the loudspeaker radiation pattern $l(f, \theta)$ for θ_l direction, the directivity pattern of the microphone $m(f, \theta)$ for θ_m, and the travelled distance d.

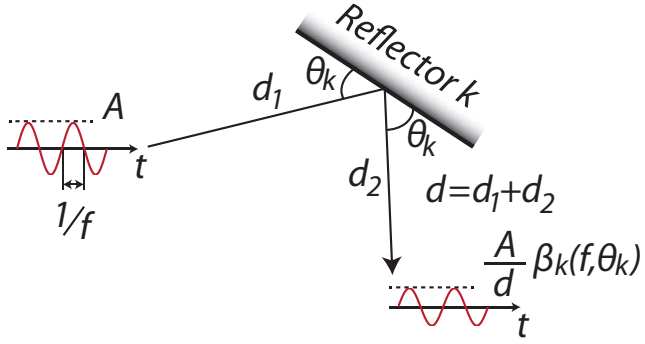


Figure 6.1.: Acoustic wave reflection from a wall.

Since the reflection coefficient values are generally frequency dependent, we estimate their values in different frequency bands. However, in order to simplify the notation, we omit in the following the subband index f. In addition, 2D wave propagation is assumed in this Chapter for clarity; an extension to 3D is straightforward and some considerations are made at the end of the Chapter.

We further assume that the source signal $s(t)$, environment geometry, $g_k(\theta)$,$l(\theta)$ and $m(\theta)$ are known *a priori*, and can be used as input to the estimation procedure.

In particular, the latter two can also be estimated or approximated using methods presented in [108], [109]. Based on this information, the parameters: θ_k, θ_l, θ_m, d, and τ (i.e., delay corresponding to the travelled distance) can be obtained from the acoustic propagation modeling engine described in Chapter 3. Thus the remaining unknowns are the signal amplitude A and the reflection coefficients of all reflective surfaces in the environment α_k.

In the following, we denote the signal acquired by microphone j as $\hat{x}_j(t)$ and the M signals acquired by the microphone array as $\hat{\mathbf{X}} = [\hat{x}_1(t), ..., \hat{x}_M(t)]^T$. On the other hand, using the *a priori* information, the signal $x_j(t)$ acquired by microphone j can be modeled. The signals x_j, for $j = 1, 2, ..., M$ at time t are organized in the vector $\mathbf{X} = [x_1(t), ..., x_M(t)]^T$. The pseudospectrum acquired by the microphone array is denoted as $\hat{P}(\theta)$. With the above a priori information at hand it is possible to model the propagation in the environment and therefore to emulate the measurement of the pseudospectrum. In particular, the model pseudospectrum is denoted by $P(\theta)$.

The aim of the estimation procedure is to match the pseudospectra $\hat{P}(\theta)$ and $P(\theta)$ for all look directions θ in order to estimate the reflection coefficients of all the walls in the acoustic enclosure.

6.4. Measurement of the pseudospectrum

When a microphone array is applied to sample the soundfield, the angular distribution of the impinging energy can be estimated using beamforming techniques and measured microphone signals. For this purpose, the room is scanned using a beamformer and the output power for each look direction forms the so-called angular pseudospectrum [76]. Such a spatial power pseudospectrum $\hat{\mathbf{P}} = [\hat{P}(\theta_1), \hat{P}(\theta_2), ...]^T$ for sample look directions $\theta_1, \theta_2, ...$ can generally be expressed as

$$\hat{P}(\theta) = \mathbf{h}^H(\theta)\hat{\mathbf{R}}\mathbf{h}(\theta), \tag{6.2}$$

where θ is the look direction, $\mathbf{h}(\theta)$ denotes the array weight vector, and $\hat{\mathbf{R}}$ denotes the estimate of the autocorrelation matrix of the microphone signals

$$\mathbf{R} = E[\hat{\mathbf{X}}\hat{\mathbf{X}}^T], \tag{6.3}$$

the signals of an M-element microphone array are given by

$$\hat{\mathbf{X}} = [\hat{x}_1(t), ..., \hat{x}_M(t)]^T,$$

where $\hat{x}_j(t)$ is the signal acquired by the jth microphone.

In general, the pseudospectrum $\hat{\mathbf{P}}$ can be measured using (6.2) and any beamforming technique. In particular, in the experiments we will consider the delay-and-sum (DAS) beamformer [76], [77] (see also the Appendix C).

6.5. Pseudospectrum Modeling

In order to model the pseudospectrum for a given acoustic scenario, the following acoustic propagation model is assumed. As illustrated in Figure 6.2, when a source signal $s(t)$ with amplitude A, emitted through a loudspeaker, reflects from a boundary in an acoustic enclosure, the incident wave is specularly reflected. The amplitude sensed by a microphone upon one reflection can then be written as [105]

$$\frac{A\alpha_k g_k(\theta_k) l(\theta_l) m(\theta_m)}{d} s(t-\tau),$$

where α_k denotes the reflection coefficient of the reflector k, $g_k(\theta_k)$ models the variation of the reflection coefficient with the angle θ_k, $l(\theta_l)$ is the loudspeaker radiation pattern, $m(\theta_m)$ is the directivity pattern of the microphone, d and τ denote the travelled distance and the corresponding delay, respectively. As stated previously, the signal $s(t)$, environment geometry, $g_k(\theta_k)$, $l(\theta_l)$ and $m(\theta_m)$ are assumed to be known, and the acoustic propagation modeling engine presented in Chapter 3 is used to obtain θ_k, θ_l, θ_m, d, and τ. Thus we can denote the known part of the signal as

$$y(t) = g_k(\theta_k) l(\theta_l) m(\theta_m) s(t-\tau)/d. \tag{6.4}$$

On the other hand, the unknowns are the reflection coefficients and the signal amplitude, which can be jointly denoted as $\Lambda = A\alpha_k$.

Upon several reflections from the walls, such as those depicted in Figure 6.3, the signal that is acquired by microphone j can be written as

$$x_j(t) = \sum_{r=0}^{R} \sum_{i=1}^{N_r} \Lambda_{ri} y_{j,ri}(t) + w_j(t), \tag{6.5}$$

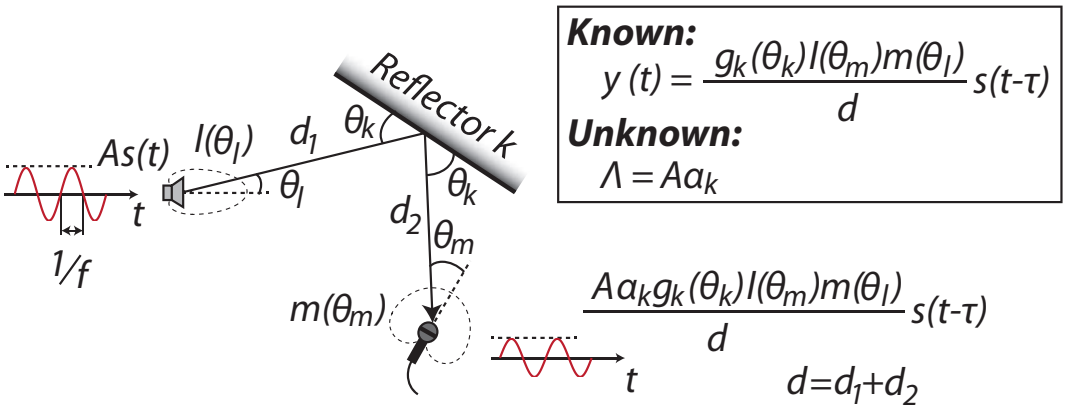


Figure 6.2.: Acoustic propagation and reflection model.

where $w_j(t)$ is the additive noise related to the self-noise of the microphone and on the AD circuitry, $y_{j,ri}(t)$ is the signal of the acoustic path with indices r and i acquired by the microphone j and r denotes the reflection order. In particular, $r = 0$ denotes the direct path, $r = 1$ the first reflection, etc.; and N_r is the number of paths for reflection order r. As an example, for a two-dimensional model of a rectangular room, $N_r = 1$ for $r = 0$ and $N_r = 4$ for $r = 1$. Λ_{ri} is the unknown amplitude of the reflected signal. As an example, $\Lambda_{01} = A$ for the direct path, $\Lambda_{1k} = A\alpha_k, k = 1, ..., N_1$ for the first-order reflection, and $\Lambda_{2i} = A\alpha_k\alpha_n, k \neq n$ for the second-order reflections, as shown Figure 6.3. We assume the additive noise $w_j(t)$ to be Gaussian, statistically independent from the signal $s(t)$ and spatially white. Early room reflections up to the order R are modeled as distinct reflections, whereas the reflection paths with $r > R$ are assumed to have very low energy and to impinge on the array uniformly from all directions, and therefore can be considered as part of the noise.

The modeled microphone signals of the M-element microphone array are written in matrix form as

$$\mathbf{X} = [x_1(t), ..., x_M(t)]^T = \sum_{r=0}^{R} \sum_{i=1}^{N_r} \Lambda_{ri} \mathbf{Y}_{ri} + \mathbf{W}, \tag{6.6}$$

where

$$\mathbf{Y}_{ri} = [y_{1,ri}(t), ..., y_{M,ri}(t)]^T,$$
$$\mathbf{W} = [w_1(t), ..., w_M(t)]^T.$$

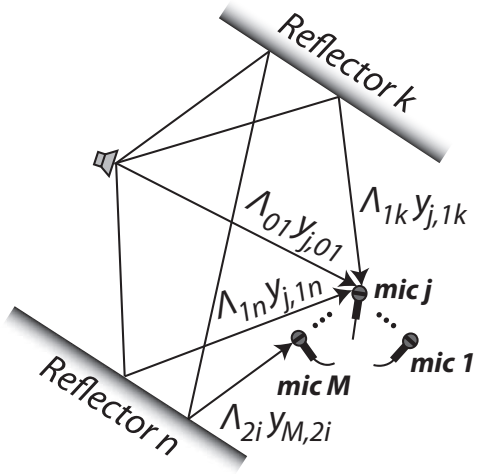


Figure 6.3.: Reflection model for an acoustic environment with two reflectors.

In the following paragraphs we put the model in (6.6) at use in modeling the pseudospectrum. Notice that the acoustic echoes coming from different paths consist of delayed and attenuated replicas of the signal $s(t)$, and are thus mutually correlated. This fact is relevant when the correlation matrix in (6.3) is modeled. In particular we indicate with $\hat{\mathbf{R}}_{r_1 i_1 r_2 i_2}$ the estimate of the correlation matrix between the generic echoes $\mathbf{Y}_{r_1 i_1}$ and $\mathbf{Y}_{r_2 i_2}$

$$\mathbf{R}_{r_1 i_1 r_2 i_2} = E[\mathbf{Y}_{r_1 i_1} \mathbf{Y}_{r_2 i_2}^T]. \tag{6.7}$$

If we use (6.7) in (6.2)and write explicitly the unknown amplitudes of different paths, we obtain

$$P(\theta) = \mathbf{h}^H (\sum_{r_1=0}^{R} \sum_{i_1=1}^{N_{r_1}} \sum_{r_2=0}^{R} \sum_{i_2=1}^{N_{r_2}} \Lambda_{r_1 i_1} \Lambda_{r_2 i_2} \hat{\mathbf{R}}_{r_1 i_1 r_2 i_2}) \mathbf{h} + \sigma_w^2 \mathbf{h}^H \mathbf{h}, \tag{6.8}$$

where $\mathbf{h}^H$ denotes the array weight vector of the respective beamformer and the noise $\mathbf{W}$ is assumed statistically independent from $\mathbf{Y}_{ri}$ and spatially white with unknown variance σ_w^2, i.e.

$$E[\mathbf{W}\mathbf{W}^H] = \sigma_w^2 \mathbf{I}_M. \tag{6.9}$$

When the DAS beamformer is in use $\mathbf{h}(\theta) = \mathbf{a}(\theta)$ and therefore we easily conclude that $\sigma_w^2 \mathbf{h}(\theta)^H \mathbf{h}(\theta) = \sigma_w^2$.

In the following, we denote with $\mathbf{P}_{r_1 i_1 r_2 i_2} = [P_{r_1 i_1 r_2 i_2}(\theta_1), ..., P_{r_1 i_1 r_2 i_2}(\theta_N)]^T$ the

contribution to the pseudospectrum coming from the acoustic paths with indexes $r1\,i1$ and $r2\,i2$, where the component $P_{r_1 i_1 r_2 i_2}(\theta)$ is

$$P_{r_1 i_1 r_2 i_2}(\theta) = \mathbf{h}(\theta)^H \hat{\mathbf{R}}_{r_1 i_1 r_2 i_2} \mathbf{h}(\theta). \tag{6.10}$$

The model of the pseudospectrum $\mathbf{P} = [P(\theta_1), \ldots, P(\theta_N)]^T$ is obtained using (6.10) in (6.8) to give

$$\mathbf{P} = \sum_{r_1=0}^{R} \sum_{i_1=1}^{N_{r_1}} \sum_{r_2=0}^{R} \sum_{i_2=1}^{N_{r_2}} \Lambda_{r_1 i_1} \Lambda_{r_2 i_2} \mathbf{P}_{r_1 i_1 r_2 i_2} + \sigma_w^2 \mathbf{v}, \tag{6.11}$$

where $\mathbf{v} = \left[\mathbf{h}(\theta_1)^H \mathbf{h}(\theta_1), ..., \mathbf{h}(\theta_N)^H \mathbf{h}(\theta_N)\right]^T$.

In order to derive a simplified expression of $\mathbf{P}$, we note that (6.11) there are contributions $\mathbf{P}_{r_1 i_1 r_2 i_2}$ whose unknown scaling factors $\Lambda_{r_1 i_1} \Lambda_{r_2 i_2}$ correspond to the same combination of reflection coefficients α_k and signal amplitude A. To clarify this fact, Figure 6.4 shows two exemplary second-order reflective paths that undergo reflections from the same walls but in different order. The angular variation of reflection coefficients is included in the model $\mathbf{Y}_{ri}$, and therefore these two paths have the same unknown amplitudes $\Lambda_{21} = \Lambda_{22} = A\alpha_k\alpha_n$. As a consequence, the autocorrelation of the first path $\mathbf{P}_{2121}$, the autocorrelation of the second path $\mathbf{P}_{2222}$, and the crosscorrelations between these two paths (namely, $\mathbf{P}_{2122}$ and $\mathbf{P}_{2221}$), all have the same unknown scaling factor $A^2\alpha_k^2\alpha_n^2$ in (6.11). We therefore group together all pseudospectrum components that have the same scaling factors. Considering the example shown in Figure 6.4, we can group these second-order reflections to obtain their weights and pseudospectrum components as $q_i = A^2\alpha_k^2\alpha_n^2$ and $\mathbf{P}_i = \mathbf{P}_{2121} + \mathbf{P}_{2222} + \mathbf{P}_{2122} + \mathbf{P}_{2221}$. After this grouping, the pseudospectrum finally takes the expression

$$\mathbf{P} = \sum_{i=1}^{N} \mathbf{P}_i q_i + \mathbf{v}\sigma_w^2 = \mathbf{M}\mathbf{q}, \tag{6.12}$$

where N is the number of pseudospectrum components after grouping, q_i denotes the scale factor of the corresponding pseudospectrum component $\mathbf{P}_i$, the model matrix $\mathbf{M} = [\mathbf{P}_1, ..., \mathbf{P}_i, ..., \mathbf{P}_N, \mathbf{v}]$, and the vector of unknowns $\mathbf{q} = [q_1, ..., q_i, ..., q_N, \sigma_w^2]^T$. As an example, $q_i = A^2$ for the autocorrelation of the direct path, $q_i = A^2\alpha_k^2$ for the autocorrelation of the first-order reflection, $q_i = A^2\alpha_k$ for the crosscorrelation between the direct path and the first-order reflection, and $q_i = A^2\alpha_k\alpha_n$ for the

crosscorrelation between two first-order reflections.

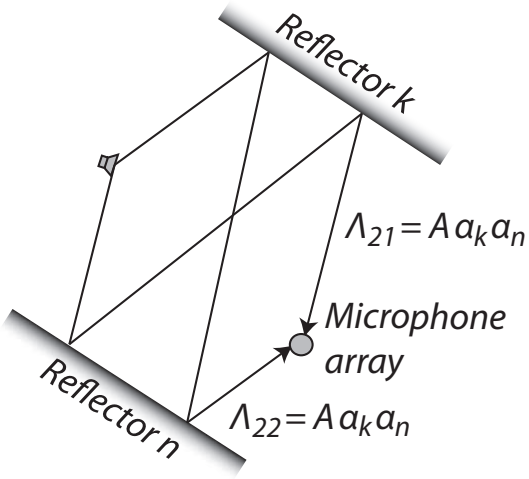


Figure 6.4.: Illustration of the amplitudes of two second-order reflections impinging on a small circular array.

Note that the scaling factors q_i are a product of a variable number of reflection coefficients α_k, of the signal amplitude A, and of the noise variance ($q_{N+1} = \sigma_w^2$). In order to simplify the derivation of the reflection coefficients from the estimates of q_i, notice that the dependency is linear if we apply a logarithmic transformation to q_i. As an example, if $q_i = A^2\alpha_k\alpha_n$, we obtain $\log(q_i) = 2\log(A) + \log\alpha_k + \log\alpha_n$. Let $\mathbf{m} = [\log A, \log\alpha_1, ..., \log\alpha_k, ..., \log\alpha_n, ..., \log\sigma_w^2]^T$ be a vector that contains logarithms of unknown parameters and let $\mathbf{H}$ be the matrix that counts the number of occurrences of these variables for each of the components of $\mathbf{q}$. As an example, the ith row of $\mathbf{H}$ corresponding to $\log q_i = \log(A^2\alpha_1\alpha_2)$ is $[2, 1, 1, 0, ..., 0]$. We can then write the relation between the component scale factors $\mathbf{q}$ and unknown parameters $\mathbf{m}$ as

$$\mathbf{d} = \mathbf{H}\mathbf{m}, \tag{6.13}$$

where $\mathbf{d} = \log(\mathbf{q})$.

6.6. Estimation

An overview of the estimation algorithm is shown in Figure 6.5. The estimation procedure is based on the matching between the measured (6.2) and simulated (6.12) pseudospectra

$$\hat{\mathbf{P}} \approx \mathbf{M}\mathbf{q}. \tag{6.14}$$

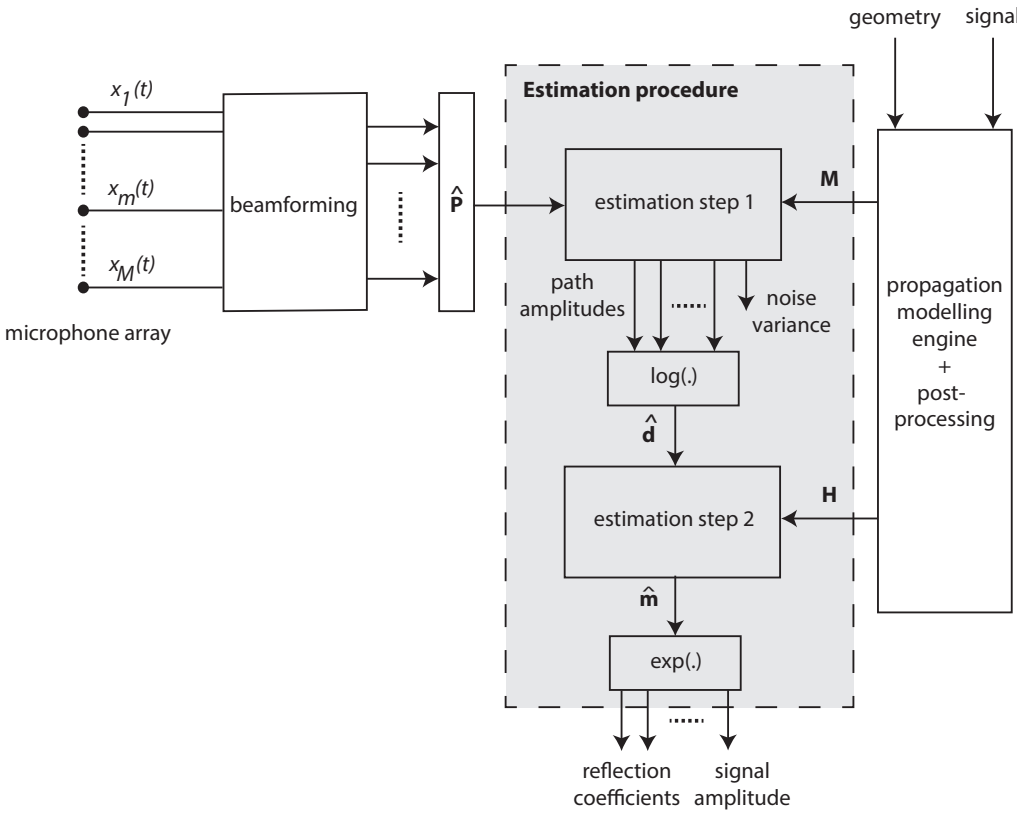


Figure 6.5.: Algorithm overview.

In (6.14), the vector $\hat{\mathbf{P}}$ represents the measurements acquired by the microphone array and $\mathbf{M}$ contains the modeled components of the pseudospectrum. The estimation is performed in two separate steps. The first step estimates the scaling factors $\mathbf{q}$. The second step estimates the reflection coefficients α_k and the signal amplitude A using the estimates $\hat{\mathbf{q}}$.

Step 1: Estimate $\mathbf{q}$ from $\hat{\mathbf{P}}$ by minimizing the sum of squared differences between the observations and the model in (6.14) (Figure 6.6). In particular the solution is obtained by adopting the constrained minimization

$$\begin{aligned} \hat{\mathbf{q}} &= \arg\min_{\mathbf{q}} \{(\hat{\mathbf{P}} - \mathbf{M}\mathbf{q})^T (\hat{\mathbf{P}} - \mathbf{M}\mathbf{q})\} \\ s.t. \quad & \mathbf{q} \geq 0, \quad q_j = q_i \alpha_k \leq q_i, \end{aligned} \tag{6.15}$$

where the constraints are used to ensure that the solutions $\hat{\mathbf{q}}$ are feasible. In par-

ticular, the constraint $\mathbf{q} \geq 0$ ensures that the scaling factors are positive, and the constraint $q_j = q_i \alpha_k \leq q_i$ ensures that the scaling factors are decreasing with the reflection order.

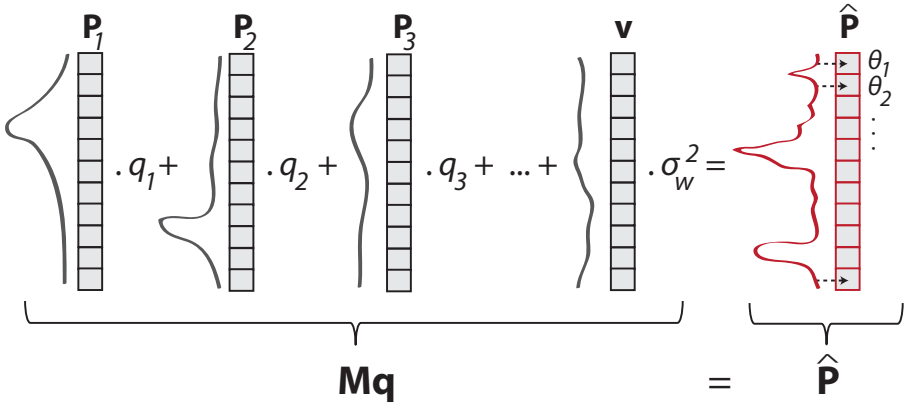


Figure 6.6.: Pseudospectrum generation model and matching with measurements.

Step 2: Obtain a least-squares (LS) solution $\hat{\mathbf{m}}$ from $\hat{\mathbf{q}}$ imposing the relation (6.13) (Figure 6.7)

$$\hat{\mathbf{m}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T \log(\hat{\mathbf{q}}). \tag{6.16}$$

Finally the estimates $[\hat{A}, \hat{\alpha}_1, \hat{\alpha}_2, ...]^T$ are obtained as $\exp(\hat{\mathbf{m}})$.

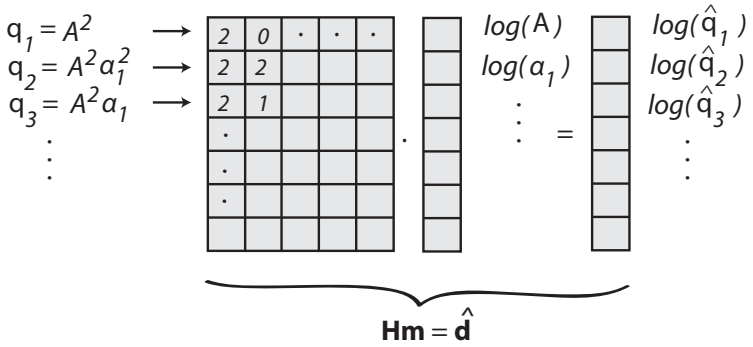


Figure 6.7.: Relation between the estimated scaling factors and unknown reflection coefficients.

Note that the number of variables $\mathbf{q}$ is much higher than the number of independent parameters to be estimated, which are signal amplitude A, reflection coefficients α_k of each reflective surface and noise variance σ_w^2. However, the two steps are performed separately and the solution $\hat{\mathbf{m}}$ is not necessary the globally best solution to

the problem. The scaling factors q_i are treated as independent in the first step (6.15) and their relationship is considered only successively in the second step (6.16). As a consequence the best solution $\hat{\mathbf{q}}$ to the first step is not necessary the one that gives the best solution $\hat{\mathbf{m}}$ to the second step. In fact, the system (6.14) is generally underdetermined or ill-conditioned which means that there could be infinite solutions $\hat{\mathbf{q}}$ that give the same pseudospectrum matching. Oversampling the pseudospectrum $\hat{\mathbf{P}}$ (i.e. increasing the sampling step $\Delta\theta$ between the pseudospectrum samples $\hat{P}(\theta_i)$ and $\hat{P}(\theta_{i+1})$) at certain point does not produce new information. This problem prevents the modeling of higher order reflections for a better matching as the number of scaling factors $\mathbf{q}$ increases considerably with the reflection order. This problem is more pronounced at lower frequencies and for absorptive materials as in such cases the pseudospectrum components $\mathbf{P}_i$ become quite smooth making $\mathbf{M}$ ill-conditioned. Different approaches can be used in order to address this issue:

- Here we first resort to a technique that keeps the same number of variables, but adds more data by combining information from different observations, i.e. building

$$\widetilde{\mathbf{M}}^T = [\mathbf{M}_{o_1}^T, \mathbf{M}_{o_2}^T, ...], \ \widetilde{\mathbf{P}}^T = [\hat{\mathbf{P}}_{o_1}^T, \hat{\mathbf{P}}_{o_2}^T, ...], \tag{6.17}$$

  where $\mathbf{M}_{o_i}$ and $\hat{\mathbf{P}}_{o_1}$ are, respectively, the model matrix and the measurement vector for the ith observation, and $\widetilde{\mathbf{M}}$ and $\widetilde{\mathbf{P}}$ the global model matrix and the global measurement vector. In experiments, multiple observations are collected by rotating the loudspeaker while keeping it fixed in the same position. This way, we probe the environment "illuminating" from different look directions, each yielding new information;

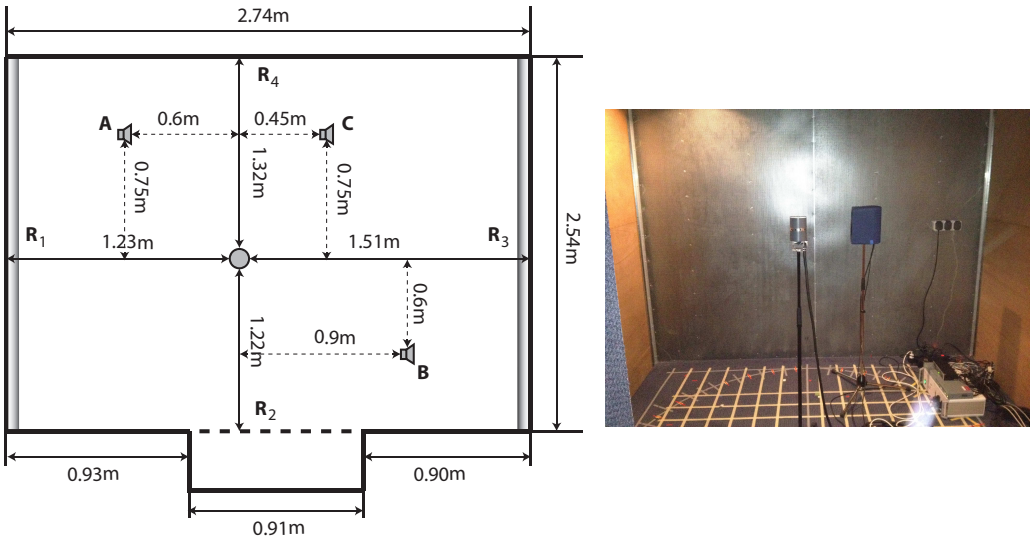
- Successively we develop an Expectation-Maximization (EM) algorithm that cycles between two estimation steps in order to improve the estimation accuracy and robustness.

6.7. Experimental evaluation

6.7.1. Setup

To verify the performance of the presented estimation algorithm, measurements were conducted in the semi-anechoic room depicted in Figure 6.8 at the Chair of Multimedia Communications and Signal Processing at Erlangen University. Two

large wooden panels are placed in front of two walls, denoted with symbols $\mathbf{R}_1$ and $\mathbf{R}_3$. The soundfield is captured with a circular microphone array that consists of ten omnidirectional microphones mounted into a rigid cylindrical baffle with a radius of 0.04 m.



(a) (b)

Figure 6.8.: Experimental setup.

In order to test the robustness of the solution the estimation is performed for three different loudspeaker positions (denoted as **A**, **B**, and **C**). For each position a different and independent estimate is obtained. By comparing the estimates we test the accuracy. A frequency constant cardioid radiation model was used to account for the the loudspeaker directivity $l(\theta)$. For each position of the loudspeaker, three different orientations have been used in order to gather multiple observations from the same source position, as required for the minimization in (6.17).

Reflections up to the second order are simulated by the modeling engine, i.e. $R = 2$. The reflection coefficients are estimated over 15 frequency bands, in particular the frequency axis is organized in fifteen frequency bands ranging from 2750 to 9750 Hz. All estimations are performed at frequencies sufficiently high so that a minimum directivity of the DAS beamformer is guaranteed. In order to obtain the estimates at lower frequencies a more robust method and a beamformer with higher directivity should be used. For the EM implementation and results see the following section.

6.7.2. Pseudospectrum matching

We first focus on the matching of the pseudospectra. Figure 6.9 shows modeled and measured pseudospectra for position **B** (with three different orientations) at 3.75kHz and 7.25kHz.

The continuous line represents the measured pseudospectrum at the given frequency band. It is obtained as described in (6.2) using the delay-and-sum (DAS) beamforming technique (see the Appendix C) applied to the signals acquired by a circular microphone array, which represents an acoustic camera that captures rays (reflections) coming from different directions.

The dashed line is obtained matching the modeled components of the pseudospectrum in (6.12), determined using the modeling engine, with the measured pseudospectrum (6.2). The matching is performed as described in (6.14) and illustrated in Figure 6.6. Therefore the dashed line corresponds to the sum of modeled components, combined using the estimated scaling factors, i.e. it corresponds to $\mathbf{M}\hat{\mathbf{q}}$, where $\hat{\mathbf{q}}$ are given by (6.15).

Notice that a good match is obtained between the modeled and measured pseudospectra, especially in proximity of the largest peaks.

6.7.3. Reflection coefficient estimates

The reflection coefficients are obtained as described in (6.16) using the scaling factors estimated from the pseudospectrum matching. The estimated reflection coefficients for each frequency band and for each wall are shown in Figure 6.10. The mean values across the frequency bands are given in Table 6.1.

Table 6.1.: Mean values of estimated reflection coefficients for different loudspeaker positions.

	$\mathbf{R}_1$	$\mathbf{R}_2$	$\mathbf{R}_3$	$\mathbf{R}_4$
Position A	0.9664	0.0795	0.9512	0.0491
Position B	0.9410	0.1351	0.9606	0.0672
Position C	0.9529	0.3641	0.9655	0.0010

The algorithm is able to estimate four reflections coefficients, all assumed to be distinct, using even a single loudspeaker position. As it can be observed, the estimates obtained for different positions are reasonably similar. Furthermore, the presence of both highly reflective and absorptive surfaces (wooden panels and the walls of the semi-anechoic room, respectively) is clearly evident. The estimated val-

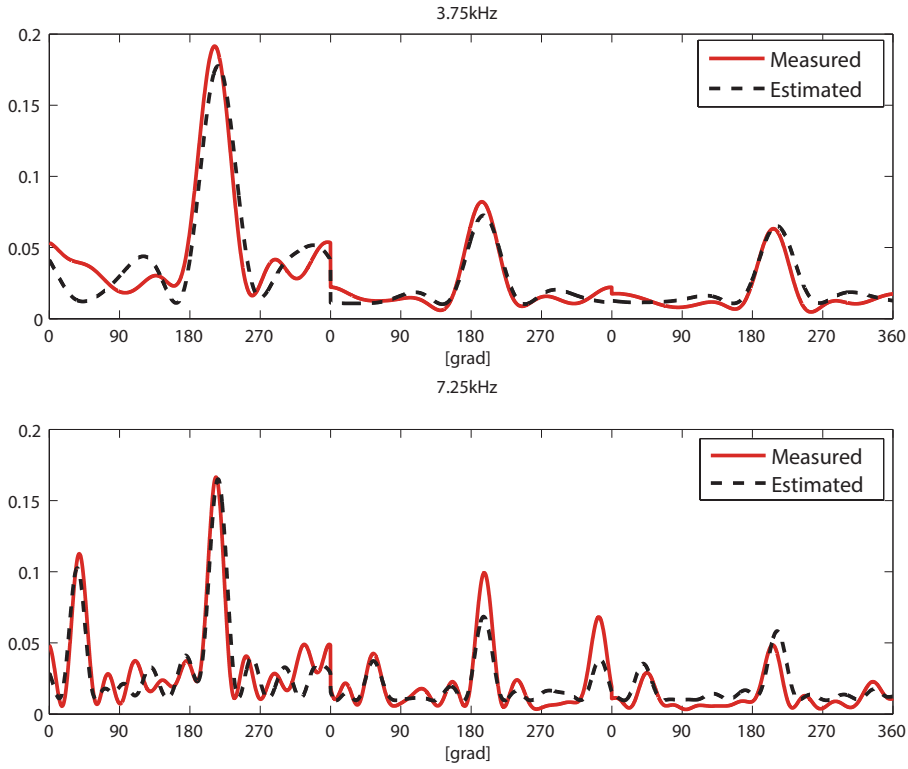


Figure 6.9.: Matching of the pseudospectra for position **B** (with three different orientations) at 3.75kHz and 7.25kHz.

ues for two wooden panels can be considered equal, and they are close to the typical values that can be found in the literature [105]; these reference values are given in Table 6.2.

Table 6.2.: Typical reflection coefficients for wooden panels.

Freq. [*Hz*]	125	250	500	1000	2000	4000
α	0.77	0.89	0.94	0.96	0.97	0.97

The higher variation of estimates of the reflectivity of $\mathbf{R}_2$ and $\mathbf{R}_4$ can be justified by the fact that the pseudospectra at secondary peaks do not match as in proximity of large ones, thus introducing a bias in the estimate. The higher variance is also explained by the logarithmic transform between estimation steps 1 and 2, which amplifies the noise/error for lower values of q_i. Furthermore, the room door is not modeled, assuming only the presence of the flat wall $\mathbf{R}_2$, and such a model deviation strongly influences the estimates (most clearly pronounced for position

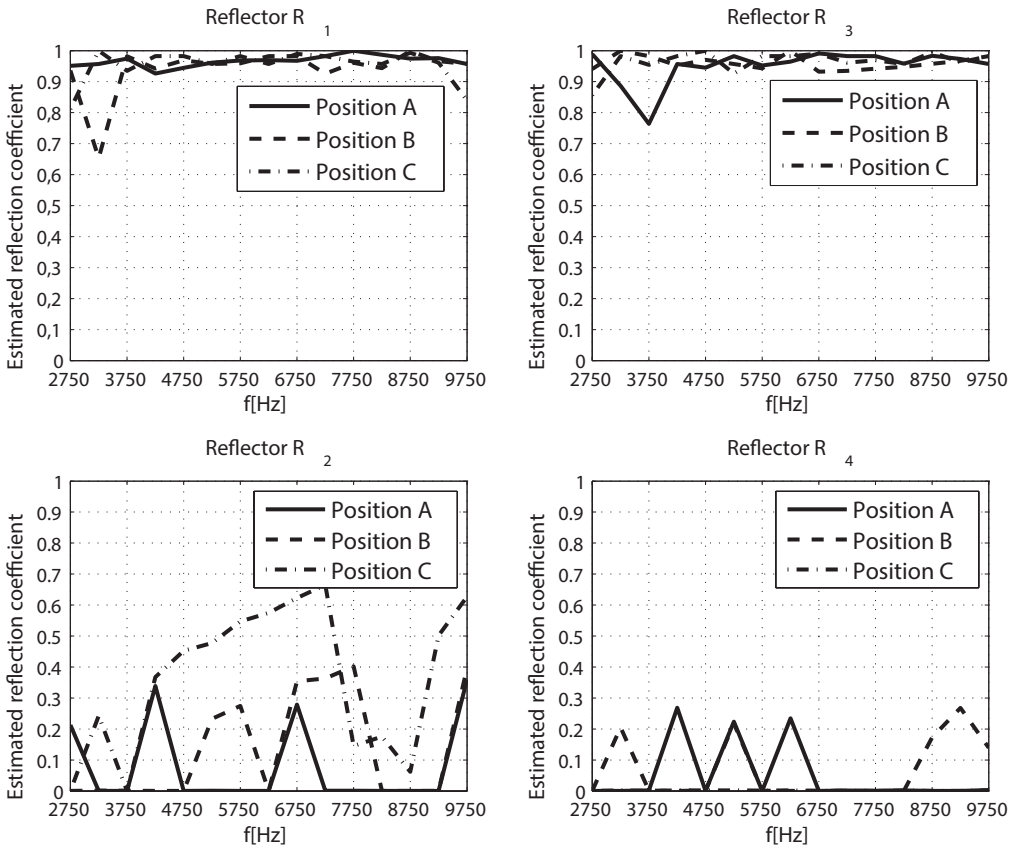


Figure 6.10.: Reflection coefficient estimates.

C). Nonetheless the lower values are in accordance with the semi-anechoic nature of the room. Finally, it is worth noting that in many typical application scenarios walls exhibit a reflection coefficient that is close to that of wood panels and will less often be close to zero.

6.8. EM algorithm

In order to improve the accuracy of the algorithm it is possible to conceive the whole estimation process as an iterative cycle between two estimation steps. To do so we treat $\mathbf{q}$ as the unobserved latent variables linked to the observations $\hat{\mathbf{P}}$ through the non-invertible transformation $\mathbf{M}$ and apply the Expectation-Maximization (EM) [110] algorithm. The EM algorithm cycles between:

1. E-step: the computation of the expected value of log-likelihood function $\log p(\mathbf{q}|\mathbf{m})$ given the observations $\hat{\mathbf{P}}$ and the solution $\hat{\mathbf{m}}^{(n)}$ at current iteration. This step performs the minimum mean square error (MMSE) [92] estimation of $\mathbf{q}$;

2. M-step: find the new solution $\hat{\mathbf{m}}^{(n+1)}$ that maximizes the estimate of log-likelihood function $\log p(\mathbf{q}|\mathbf{m})$ obtained in the previous step.

Assuming the pseudospectrum noise $\mathbf{n}$ and component scale factors $\mathbf{q}$ normally distributed with covariances $\mathbf{C}_{nn}$ and $\mathbf{C}_{qq}$ respectively, we have

$$\hat{\mathbf{P}} = \mathbf{M}\mathbf{q} + \mathbf{n},$$

$$\mathbf{n} \sim N(\mathbf{0}, \mathbf{C}_{nn}),\ \mathbf{q} \sim N(e^{\mathbf{H}\mathbf{m}}, \mathbf{C}_{qq}).$$

The E-step calculates the expected value of log-likelihood function given the current estimate $\hat{\mathbf{m}}^{(n)}$

$$Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)}) = E[\log p(\mathbf{q}|\mathbf{m})|\hat{\mathbf{P}}, \hat{\mathbf{m}}^{(n)}]. \tag{6.18}$$

For the Gaussian distribution the log-likelihood $\log p(\mathbf{q}|\mathbf{m})$ is given by

$$\begin{gathered}\log p(\mathbf{q}|\mathbf{m}) = c - \tfrac{1}{2}(\mathbf{q} - e^{\mathbf{H}\mathbf{m}})^T \mathbf{C}_{qq}^{-1}(\mathbf{q} - e^{\mathbf{H}\mathbf{m}}) \\ = \tfrac{1}{2} e^{\mathbf{H}\mathbf{m}^T} \mathbf{C}_{qq}^{-1}\mathbf{q} + \tfrac{1}{2}\mathbf{q}^T \mathbf{C}_{qq}^{-1} e^{\mathbf{H}\mathbf{m}} - \tfrac{1}{2} e^{\mathbf{H}\mathbf{m}^T} \mathbf{C}_{qq}^{-1} e^{\mathbf{H}\mathbf{m}} + d,\end{gathered}$$

where c and $d = c - \frac{1}{2}\mathbf{q}^T \mathbf{C}_{qq}^{-1}\mathbf{q}$ indicate the terms that do not depend on $\mathbf{m}$. The expected value (6.18) is then given by

$$\begin{gathered}Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)}) = \tfrac{1}{2} e^{\mathbf{H}\mathbf{m}^T} \mathbf{C}_{qq}^{-1} \hat{\mathbf{q}}(\mathbf{m}^{(n)}) + \\ \tfrac{1}{2}\hat{\mathbf{q}}(\mathbf{m}^{(n)})^T \mathbf{C}_{qq}^{-1} e^{\mathbf{H}\mathbf{m}} - \tfrac{1}{2} e^{\mathbf{H}\mathbf{m}^T} \mathbf{C}_{qq}^{-1} e^{\mathbf{H}\mathbf{m}} + d \\ = -\tfrac{1}{2}(\hat{\mathbf{q}}(\mathbf{m}^{(n)}) - e^{\mathbf{H}\mathbf{m}})^T \mathbf{C}_{qq}^{-1}(\hat{\mathbf{q}}(\mathbf{m}^{(n)}) - e^{\mathbf{H}\mathbf{m}}) + e,\end{gathered} \tag{6.19}$$

where $e = d + \frac{1}{2}\hat{\mathbf{q}}(\mathbf{m}^{(n)})^T \mathbf{C}_{qq}^{-1}\hat{\mathbf{q}}(\mathbf{m}^{(n)})$ and $\hat{\mathbf{q}}(\mathbf{m}^{(n)}) = E[\mathbf{q}|\hat{\mathbf{P}}, \mathbf{m}^{(n)}]$ is the MMSE estimate of $\mathbf{q}$. For $\hat{\mathbf{P}}$ and $\mathbf{q}$ jointly Gaussian we have the conditional probability distribution of $\mathbf{q}$ given $\hat{\mathbf{P}}$

$$\mathbf{q}|\hat{\mathbf{P}} \sim \mathcal{N}(\mu_q + \mathbf{C}_{qP}\mathbf{C}_{PP}^{-1}(\hat{\mathbf{P}} - \mu_P), \mathbf{C}_{qq} - \mathbf{C}_{qP}\mathbf{C}_{PP}^{-1}\mathbf{C}_{Pq}),$$

and thus the MMSE estimate $\hat{\mathbf{q}}(\mathbf{m}^{(n)})$ is the mean value of a posteriori probability $p(\mathbf{q}|\hat{\mathbf{P}}, \mathbf{m}^{(n)})$, i.e.

$$\hat{\mathbf{q}}(\mathbf{m}^{(n)}) = E[\mathbf{q}|\hat{\mathbf{P}}, \mathbf{m}^{(n)}] = \mu_q + \mathbf{C}_{qP}\mathbf{C}_{PP}^{-1}(\hat{\mathbf{P}} - \mu_P)$$

where

$$\mu_q = E[\mathbf{q}] = e^{\mathbf{H}\mathbf{m}^{(n)}},$$

$$\mu_P = E[\hat{\mathbf{P}}] = \mathbf{M}\mu_q,$$

$$\mathbf{C}_{PP} = E[(\hat{\mathbf{P}} - \mu_P)(\hat{\mathbf{P}} - \mu_P)^T] = \mathbf{M}\mathbf{C}_{qq}\mathbf{M}^T + \mathbf{C}_{nn},$$

$$\mathbf{C}_{qP} = E[(\mathbf{q} - \mu_q)(\hat{\mathbf{P}} - \mu_P)^T] = \mathbf{C}_{qq}\mathbf{M}^T.$$

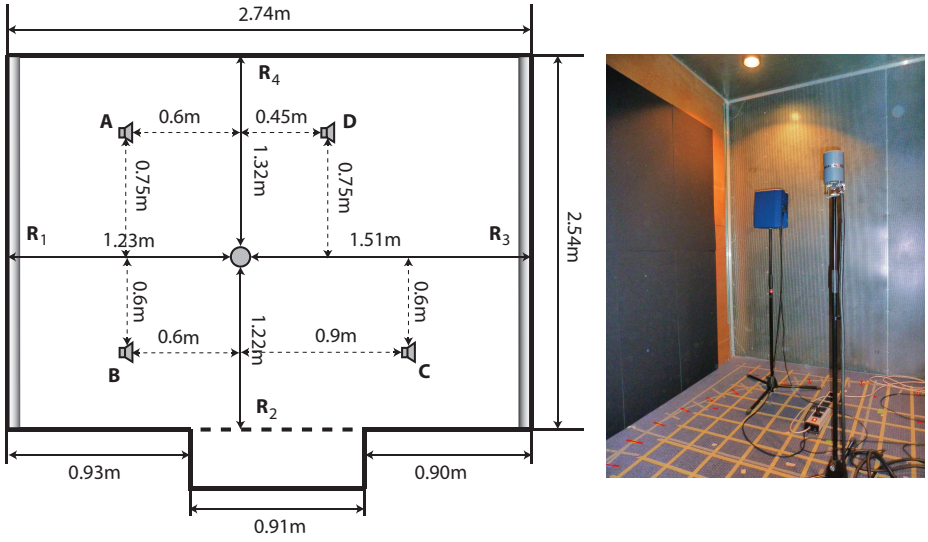
The M-step finds the new estimate $\hat{\mathbf{m}}^{(n+1)}$ maximizing the expected value of log-likelihood function $Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)})$ computed using (6.19)

$$\hat{\mathbf{m}}^{(n+1)} = \arg\max_{\mathbf{m}} Q(\mathbf{m}|\hat{\mathbf{m}}^{(n)})$$
$$= \arg\min_{\mathbf{m}}\{(\hat{\mathbf{q}}(\mathbf{m}^{(n)}) - e^{\mathbf{H}\mathbf{m}})^T\mathbf{C}_{qq}^{-1}(\hat{\mathbf{q}}(\mathbf{m}^{(n)}) - e^{\mathbf{H}\mathbf{m}})\}.$$

6.9. Experimental evaluation (EM)

The performance of the EM algorithm is tested in the same semi-anechoic room and with the same equipment as done in the previous experiment (see Figure 6.11). According to the specific experiment, walls $\mathbf{R}_1$ and $\mathbf{R}_3$ can be covered with reflective or semi-reflective materials, whereas walls $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive. Notice that this is a challenging scenario. In fact peaks related to reflective paths bouncing off absoprptive walls are dimmed, thus making the estimation of reflection coefficients of absorptive walls sensitive to even small measurement and modeling errors. Therefore, the frequency-dependent radiation pattern of GENELEC 1029A loudspeaker that probes the environment has been provided by the manufacturer. The loudspeaker, especially at high frequencies, is strongly directional and presents the maximum directivity in the frontal part. Reflectors that are located on the rear of the speaker, therefore, would not be adequately probed. In order to overcome this issue, four different orientations (0°, 90°, 180° and 270°) for each position of the speaker are adopted similarly as done in the previous experiment. Reflections up to the third order are simulated by the modeling engine, i.e. $R = 3$, and the room

door is taken into account by the modeling engine.



(a) (b)

Figure 6.11.: Experimental setup for EM algorithm.

In order to test the robustness of the solution the estimation is performed at four different loudspeaker positions, which are denoted in Figure 6.11 by symbols **A**, **B**, **C** and **D**. Again, we expect estimates to be as much as possible position independent.

In the first scenario walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with two distinct reflective materials. More specifically, $\mathbf{R}_1$ is covered with aixFOAM ® semi-reflective material (thickness 30mm [1]) and $\mathbf{R}_3$ is covered with Sonatech ® reflective material (thickness 10 mm [2]). Datasheets present reflective properties in terms of absorption coefficient a, which is related to the reflection coefficient α by

$$\alpha = \sqrt{1-a}.$$

The covariance matrices $\mathbf{C}_{nn}$ and $\mathbf{C}_{qq}$ are assumed to be diagonal and their values are chosen empirically as a trade-off between a good matching of $\mathbf{M}\hat{\mathbf{q}}$ to $\hat{\mathbf{P}}$ (smaller variance of $\mathbf{n}$) and a good matching of $e^{\mathbf{H}\hat{\mathbf{m}}}$ to $\hat{\mathbf{q}}$ (smaller variance of $\mathbf{q}$).

Estimations have been performed with both DAS and MVDR beamformers. As

[1] datasheet at `http://www.aixfoam.com/absorption-foam-sh0061-felt-lamination`
[2] `http://www.sonatech.de/konfiguration/dateienpdf/SONATECH_PUR_SKIN_ProspektSeite6.pdf`

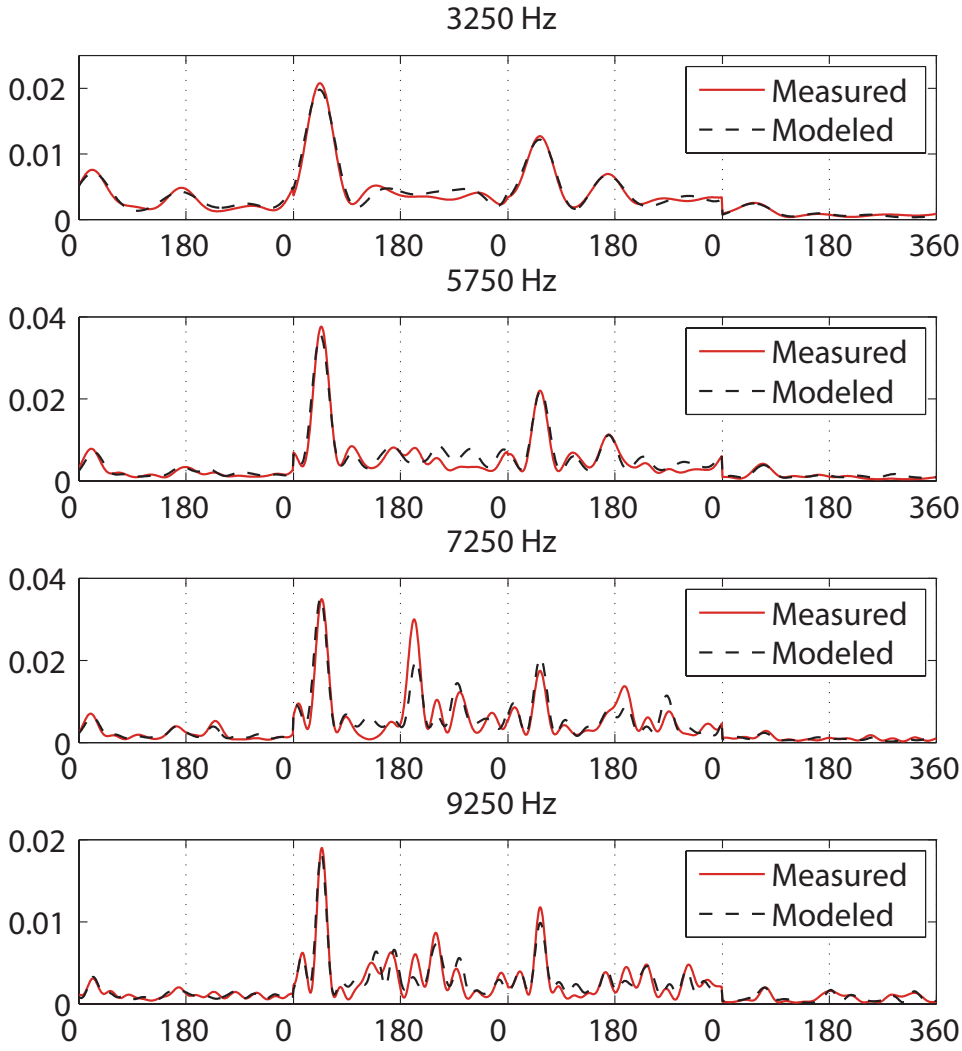


Figure 6.12.: Examples of matching between modeled and measured pseudospectra for different frequencies; Delay and Sum Beamformer is in use; position of the loudspeaker is **A** with orientations 0°, 90°, 180° and 270°.

for the DAS beamformer, it provides estimates over 20 sub-bands, whose central frequencies range from 250 to 9750 Hz. The algorithm first estimates the reflection coefficient at 9750 Hz and then recursively initializes the estimate at the lower subband with the estimate at the upper subband. The choice of starting with the highest subband is dictated by the higher resolution of the pseudospectrum at high frequencies, which enables a more accurate matching between $\hat{\mathbf{P}}$ and $\mathbf{P}$.

The pseudospectrum $\hat{\mathbf{P}}$ measured at the position $\mathbf{A}$ and the estimated pseudospec-

trum $\mathbf{M}e^{\mathbf{H}\hat{\mathbf{m}}}$ at different frequencies are shown in Figure 6.12. Notice that a good match exists, especially at high frequencies and for relevant peaks.

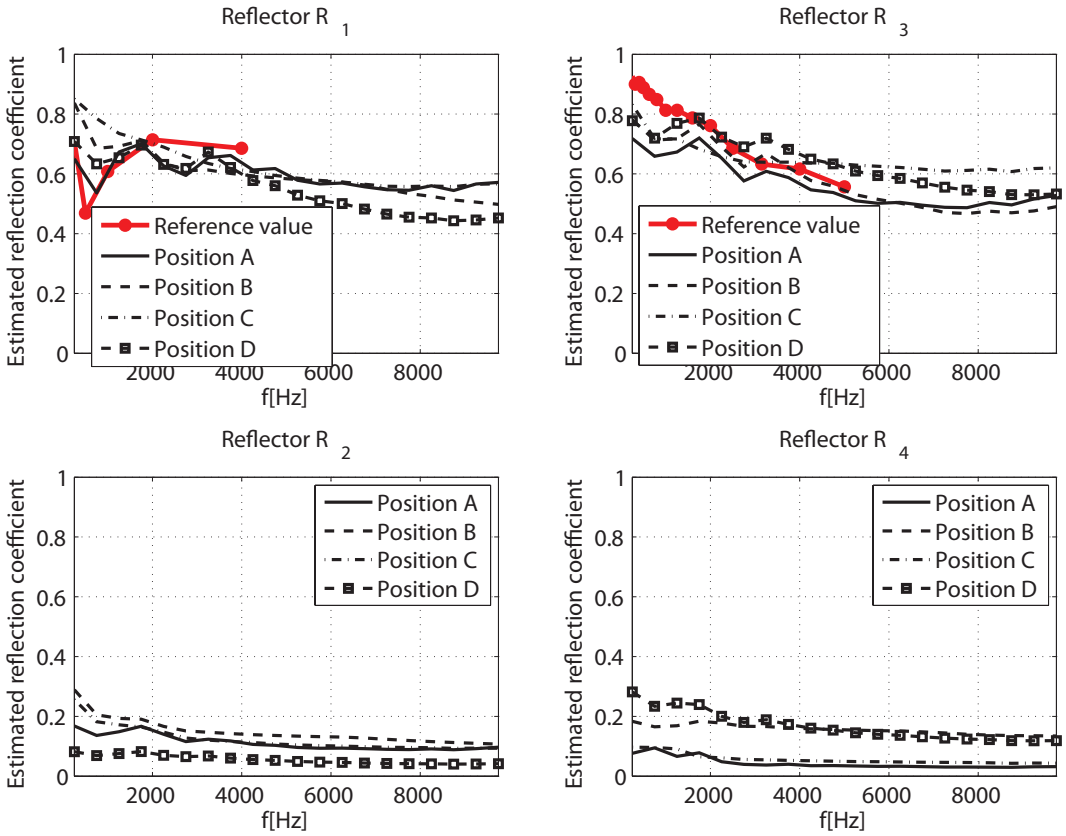


Figure 6.13.: Estimates using the DAS beamformer in the first scenario. Walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with reflective materials. Wall $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

The estimations of the frequency-dependent reflection coefficients for $\mathbf{R}_1, \ldots, \mathbf{R}_4$ along with reference values for walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are shown in Figure 6.13. The estimated values are consistent for all the loudspeaker positions. Furthermore, the estimates for walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are close to the values specified in the data-sheets, while the values for walls $\mathbf{R}_2$ and $\mathbf{R}_4$ are in accordance with the semi-anechoic nature of the room.

To further test the ability of the proposed approach to correctly estimate the reflection coefficients of different reflectors in the environment in presence of both highly reflective and absorptive surfaces we repeated the experiment in other two scenarios. In the second scenario the wall $\mathbf{R}_1$ is covered with a wooden panel and

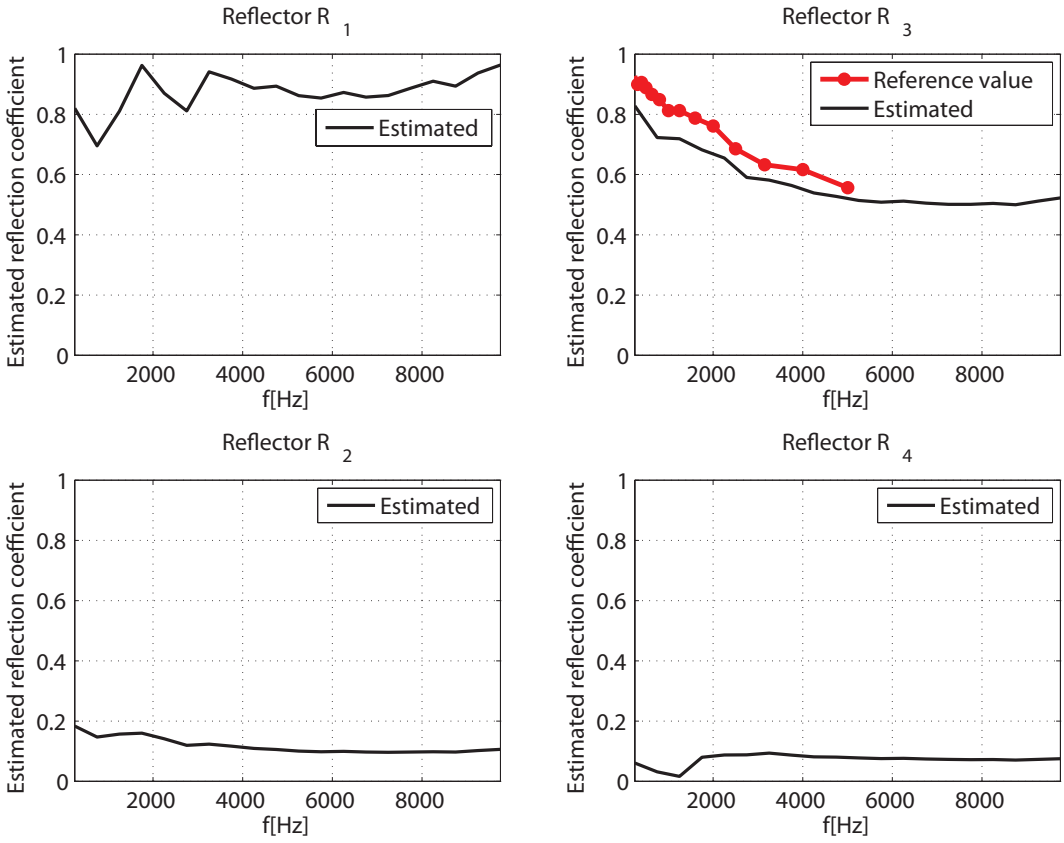


Figure 6.14.: Estimates using the DAS beamformer for position **A** in the second scenario. Wall $\mathbf{R}_1$ is covered with a wooden panel and $\mathbf{R}_3$ covered with a reflective material. Wall $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

the wall $\mathbf{R}_2$ is covered with aixFOAM ®. The estimates at position **A** are shown in Figure 6.14. As expected, the estimate of $\mathbf{R}_2$ is reasonably similar to the one obtained in the first scenario while the estimate of $\mathbf{R}_1$ has values that are close to the typical values for wooden panels found in the literature [105].

In the third scenario only the wall $\mathbf{R}_1$ is covered with aixFOAM ®, whilst all the other walls are absorptive. The estimates at the position **A** are shown in Figure 6.15. The estimate of $\mathbf{R}_1$ is similar to the estimate obtained in the first scenario, whereas the estimate of $\mathbf{R}_2$ has a low value similar to the values estimated for the other absorptive walls of room.

MVDR beamforming is used to improve the estimation at lower frequencies and for absoptive walls. Estimates are obtained over 20 frequency bands, ranging from 250 to 5000Hz. Figure 6.16 shows the reflection coefficient estimates in the first

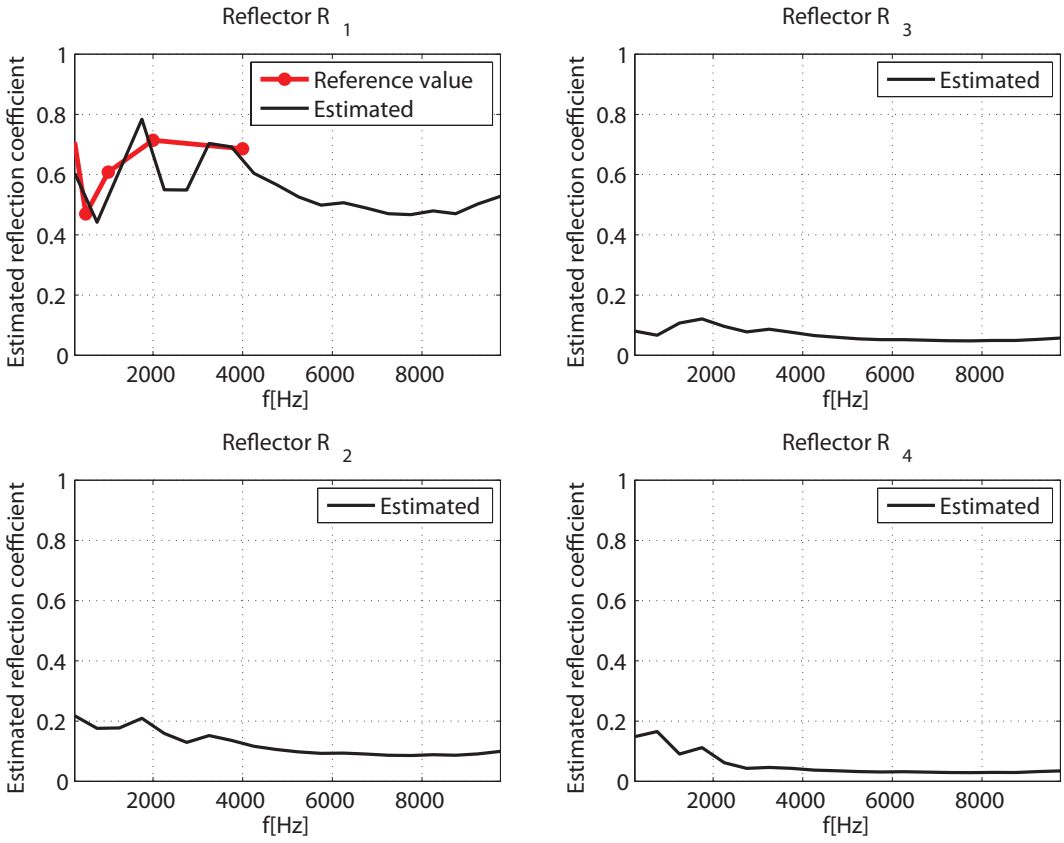


Figure 6.15.: Estimates using the DAS beamformer for position $\mathbf{A}$ in the third scenario. Wall $\mathbf{R}_1$ is covered with a reflective panel and $\mathbf{R}_3$. Walls $\mathbf{R}_2, \ldots, \mathbf{R}_4$ are absorptive.

scenario obtained using the MVDR beamformer. The results are obtained over 20 frequency bands ranging from 250 to 5000 Hz. For each band the pseudospectrum is computed using frequency smoothing on 5 frequency bins.

Notice that for all the positions and for all the walls the reflection coefficients estimated with MVDR is much closer to the reference value with respect to the estimate obtained with DAS in the same scenario, and reported in Figure 6.13. This is true especially for the absorptive walls. Notice, in fact, that reflection coefficient of walls $\mathbf{R}_2$ and $\mathbf{R}_4$ in Figure 6.16 exhibits a trend that is consistent for all the tested positions of the speaker, whereas the estimates with DAS are characterized by a greater variance among different positions.

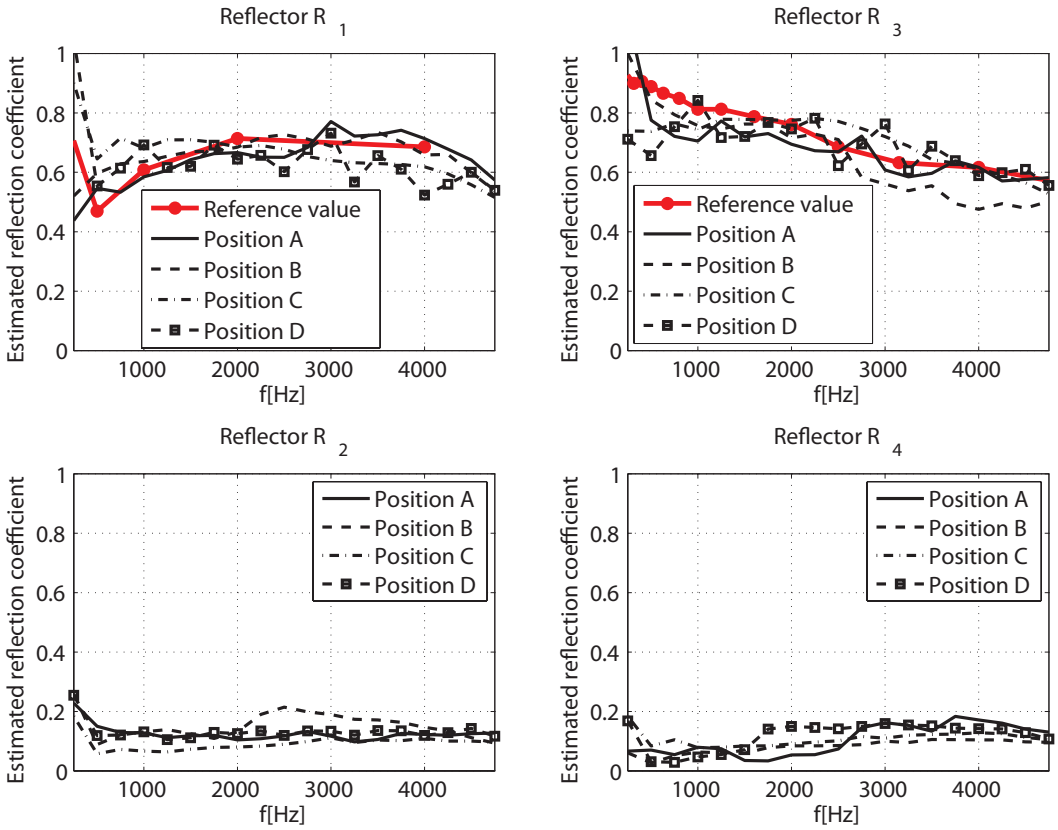


Figure 6.16.: Estimates using the MVDR beamformer in the first scenario. Walls $\mathbf{R}_1$ and $\mathbf{R}_3$ are covered with reflective materials. Wall $\mathbf{R}_2$ and $\mathbf{R}_4$ are absorptive.

6.10. Perspective work: extension to 3D

Given the 3D modeling engine presented in the Chapter 3 the extension of the presented reflection coefficient estimation technique to 3D environments is straightforward using a spherical microphone array. In a 3D geometry, the loudspeaker radiation pattern $l(\theta, \phi)$; the directivity pattern of the microphone $m(\theta, \phi)$; and the angular variation of the reflection coefficient $g_k(\theta, \phi)$, are all a function of two angles θ and ϕ that specify a direction in a 3D space. The measured pseudospectra in (6.2), obtained applying a beamforming technique to the signals acquired by the spherical microphone array, is a function of θ and ϕ as well. Therefore the pseudospectrum data $\hat{\mathbf{P}}$ has a matrix form, with rows and columns corresponding to samples of θ and ϕ respectively. The same applies to the modeled components of the pseudospectra $\mathbf{P}_i$ in (6.12) obtained using the 3D modeling engine and following the same steps

described for the 2D algorithm. The unknown scale factors $\mathbf{q}$ are still a product of a variable number of reflection coefficients α_k and the signal amplitude A (the terms $g_k(\theta, \phi)$ that specify the angular variation of the reflection coefficients are included in the model). Reordering the matrices $\hat{\mathbf{P}}$ and $\mathbf{P}_i$ into vectors (i.e. using each column as a new observation and combining them as done in (6.17)) we obtain the same formulation as in (6.14). Consequently we can apply the same estimation procedures described previously for the estimation of the reflection coefficients.

6.11. Conclusive remarks

In this Chapter a novel model-based method for the estimation of reflection coefficients has been proposed. This method matches the pseudospectrum obtained using the acoustic propagation modeling engine described in Chapter 3 with the pseudospectrum obtained using the acoustic measurements acquired by a compact microphone array, which represents a narrow version of the plenacoustic image introduced in Chapter 4. Experiments with real data show promising results and confirm the applicability of the presented method for the reflection coefficient estimation in real acoustic environments. In fact, the experimental results proved that the method is both accurate and robust: the estimated values are close to the reference values, specified either by datasheets or found in literature; and the algorithm performs consistently well for different loudspeaker positions and in different scenarios.

Chapter 7

Application to rendering – discussion

THIS Chapter brings together different aspects examined in this dissertation and shows how the outputs of different algorithms can be used together in a single application. We discuss how the beam tracing modeling engine and the knowledge about geometric and radiometric properties of the environment can be used in rendering applications. In particular, we discuss a method for rendering virtual acoustic scenes in reverberant environments. This method is based on [111], [8] and [9], but it extends its validity to the case of reverberant environment through a novel approach to room compensation. Previous solutions started from the assumption that the acoustics of the environment in which the rendering takes place ("hosting" environment) has a negligible impact on the rendering quality. This, in fact, is only reasonable if heavy acoustic conditioning is in place. Therefore, the reverberations of the hosting environment have to be taken into account. However, in order to take into account the reverberations of the hosting environment, its geometry and reflection coefficients have to be known or estimated. The beam tracing is then used for the simulation of both the virtual scene to be rendered and reverberations of the real environment to be compensated. This information is then used by the rendering engine to reproduce the desired soundfield while compensating for undesired reflections caused by the environment that hosts the system. In this Chapter we focus

on describing how the output of the beam tracing can be incorporated in the rendering system, for a detailed description of the rendering problem, theoretical background and proposed solutions see [112].

7.1. Motivations

Acoustic wavefield rendering is aimed at producing a desired soundfield in a prescribed region of space (rendering space) using a spatial distribution of loudspeakers. This is a problem of growing interest in the research community for its many applications to telecommunications (e.g. immersive telepresence), multimedia (e.g. immersive gaming, automotive, and entertainment in general), and professional audio. Some of the solutions offered in the literature are aimed at reproducing an acquired/estimated wavefield (data-driven methods). Others specify the acoustic scene as a spatial distribution of sources and acoustic reflectors (model-driven methods [113]). An approach of this sort has the advantage of flexibility, and paves the way to applications of interactive virtual acoustics (e.g. immersive gaming, walk-through architectural acoustics, etc.).

7.2. Related work

Rendering – The literature is rich with solutions for soundfield reconstruction, the most widespread of which are based on WaveField Synthesis (WFS) [114, 115]; on spherical harmonics (e.g. Higher Order Ambisonics (HOA) [116, 117]); and on the least-squares (LS) minimization of the rendering equation [118, 119]. In WFS the loudspeakers of an array act as secondary sources, in compliance with the Huygens principle, for the reconstruction of planar or spherical wavefronts and combinations thereof. HOA exploits a decomposition of the soundfield in spherical harmonic functions and performs soundfield reconstruction with a spatial distribution of loudspeakers around the listener. LS techniques are aimed at determining the filters that best reconstruct the wavefield in a number of spatially distributed control points, in compliance with a rendering equation, which is defined according to the soundfield representation. An example of rendering equation based on a spherical/circular harmonics wavefield decomposition is proposed in [120]; and one based on geometric wavefield decomposition, is discussed in [8].

Compensation – Room compensation is a well-established research topic, an

early overview of which can be found in [121]. In [122] multichannel filtering is proposed for correcting room effects in selected points within the listening area. In [123] an adaptive compensation method based on modal wavefield analysis was proposed, which extends the validity of room compensation on a wider region surrounding the control points. This method was further refined in [124], and later enriched with a concentric microphone array in [125]. A different approach is proposed in [83], where an accurate estimation of room responses over the whole rendering area is used for improving room compensation.

7.3. Problem formulation

One crucial aspect of model-driven rendering is that the image sources produced by the walls of the virtual environment "illuminate" the rendering space with an acoustic beam that is reflected and subdivided during the propagation (see Chapter 3). As a consequence of the splitting process, the beam narrows as the order of reflection grows. Therefore, the image source needs to be angularly windowed in order to account for the visibility through the chain of virtual "mirrors" that generated it. In [126, 127, 128] the image source method is used for computing the position of the image sources. A visibility test is then performed to assess the visibility between each virtual image source (an image source generated by reflecting a source to be rendered about one or more virtual walls) and each point in the listening area, which is quite a demanding task even for simple geometries [24, 14]. On the other hand, instead of being assessed in a point-to-point fashion, the visibility can be computed very efficiently and all at once through a beam tracing algorithm presented in the Chapter 3.

Beam tracing is an important tool not only for the computation of the position and visibility conditions of the image sources, but also because it offers a powerful tool for modeling the propagation of sound in enclosures. In fact, instead of directly solving the wave equation for the points of interest, beam tracing encodes the visibility of image sources with acoustic beams and models the interaction of the wavefield with the environment as the propagation, branching and reflection of beams. Therefore, the beam tracing approximates the wavefield as the superposition of the wavefields produced by image sources, each visible in a prescribed region (i.e. the area covered by the beam). Each beam is characterized by its orientation, aperture and position of the image source. This provides the specifications of the acoustic beams that

must be reconstructed and overlaid in order to form the desired wavefield. As an example, Figure 7.1 shows the superposition of beams predicted by beam tracing in a densely occluded enclosure. For clarity of visualization the figure shows only beams up to the second order of reflection.

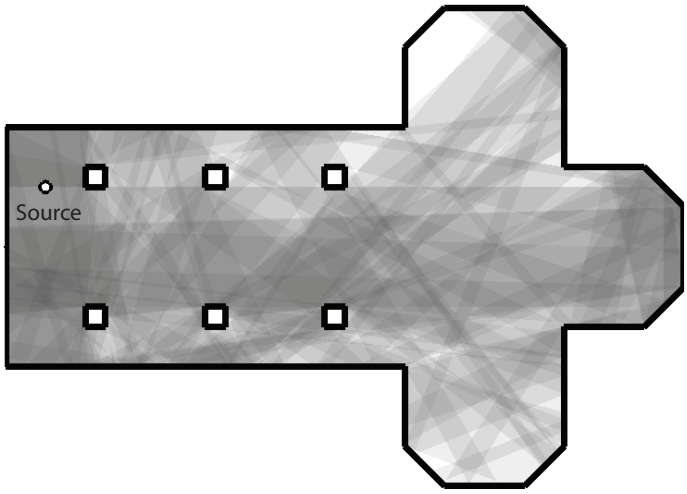


Figure 7.1.: The superposition of beams predicted by beam tracing in a densely occluded environment; beams up to the second order of reflection are shown; the figure shows the superposition of geometric areas interested by single beams – the radiometric information of a beam can be assessed by examining the beam tree, i.e. the reflection history and propagation distance.

In this work we are interested not only in reproducing the acoustics of a virtual environment using a loudspeaker array, but also in reducing the natural reverberation generated by the real environment that hosts the rendering system. In order to be taken into account, the real environment has to be modeled, once again, using the beam tracing method. This, however, means that the geometry and reflective properties of the hosting environment have to be known a priori. Alternatively, the geometry of the environment can be estimated using the algorithm described in Chapter 5, and the reflection coefficients using the algorithm described in Chapter 6. In the simulations we examine the robustness of the proposed approach to soundfield rendering with respect to modeling errors in both geometry and reflection coefficient.

7.3.1. Application scenario

The goal of a *rendering system* is to accurately reproduce a desired wavefield within a prescribed *listening area*, using a spatial distribution of *loudspeakers*. The

listening area is defined by a set of *control points*.

A simple 2D example of rendering scenario is shown in Figure 7.2, where the loudspeakers (indicated with black crosses) are placed on a circle around the listening area (gray region), which is sampled by a regular distribution of control points (black dots). In the simplest case the goal of this rendering system is to generate the wavefield corresponding to an omnidirectional virtual source (red circle) within the rendering area, under the assumption that the rendering environment (host) be anechoic.

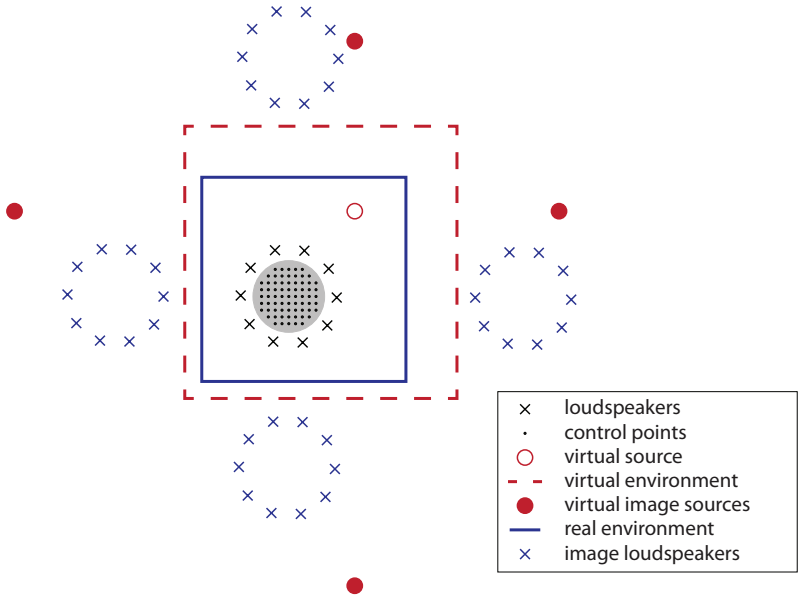


Figure 7.2.: Geometry of a rendering system operating in a real environment, aimed at reconstructing a virtual acoustic scene.

More challenging is the problem of rendering an acoustic scene in which the virtual source is surrounded by a *virtual environment*. Because of the virtual walls, in fact, we need to account for a multitude of reflective acoustic paths that form a virtual Room Impulse Response (RIR). A typical RIR exhibits two fairly distinguished portions: early reflections and late reverberations. The former is characterized by isolated distinguishable peaks, while the latter looks more like an exponential decay of densely packed reflective paths. Early reflections are known to convey information about the source position and radiance as well as the location of the reflectors, while late reverberations carry information on the size of the environment and its acoustic quality. We focus on the rendering of early reflections caused by a virtual

environment. In order to render such reflections we need to render the *virtual image sources* obtained by mirroring the virtual source on the virtual reflectors. With reference to Figure 7.2, the red dotted line denotes the contour of a rectangular virtual room (each segment of the room represents a planar reflector). The corresponding virtual image sources are denoted by red circles.

So far we have neglected the impact of the *host* environment on wavefield rendering, as we have assumed the propagation from each loudspeaker to each control point to take place in free-field conditions. Our goal, however, is to devise a methodology for compensating, at least partially, the reflections caused by the host environment on the rendered wavefield. We focus on the compensation of early reflections only. There are several reasons behind this choice. On one hand, the spatial impression induced in the listener can be mostly attributed to early reflections [129, 113]; we can also reasonably assume that the hosting environment has a high clarity index (i.e. early reflections dominate over late reverberations); finally, this choice allows us to use the same geometric tools for both room compensation and wavefield rendering in a computationally efficient fashion. If the real environment is made of planar reflectors the early reflections can be accounted for by introducing *image loudspeakers*. In Figure 7.2 the continuous blue line represents the real environment and the blue crosses denote the image loudspeakers.

It is important to notice, however, that image sources cannot be rendered with their original beam pattern, as we need to account for their visibility "through" the walls that generated such images through reflections. Therefore, computing the visibility of an image source is a necessary step for correctly accounting for wall reflections. This is true for both real sources (loudspeakers) and virtual sources, and for both real and virtual walls. The visibility is here represented by a beam, which can be interpreted as an angular windowing function acting on the original beam pattern of the source.

In this work we address the rendering problem in 2D geometries. Notice, however, that the considerations made in this paragraph have general validity.

7.3.2. Requirements

In the light of the issues introduced in the previous paragraph, we now identify the requirements of a rendering system that reproduces the acoustics of a virtual environment when it is operating in a reverberant room. We can distinguish the

following components:

1. *A propagation modeling technique that traces beams as they propagate in the virtual environment.* This operation can be efficiently performed by employing the beam tracing method presented in Chapter 3;

2. *A detailed information on the hosting environment geometry and corresponding reflection coefficients.* If this information is not known a priori the algorithms described in Chapters 5 and 6 can be employed;

3. *A propagation modeling technique that traces image loudspeakers along with their visibility conditions in the environment hosting the array.* This operation is, again, efficiently performed by the beam tracing engine using the geometric and radiometric information about the hosting environment;

4. *A beam-shaping engine that allows us to render a beam in the listening area by means of a loudspeaker array.* For more details see the Appendix D;

5. *A methodology that exploits the knowledge of the image speakers and of their visibility conditions for the compensation of the early reflections produced by the environment hosting the rendering system.* This technique is referred to as room compensation.

Let us consider the simple scenario in which a virtual source is rendered in an anechoic host environment. In this situation all we need is a beam-shaping component that approximates the source's beam pattern. If we introduce a virtual environment, the associated early reflections can be accounted for through a process of beam tracing that computes the locations of the virtual image sources along with the corresponding visibility functions (beams). Finally, if the rendering system is placed in a reverberant host environment, we need to introduce a room compensation technique for removing the unwanted early reverberations caused by the real walls. In this case, all the above components are jointly used to achieve the desired rendering result and the corresponding global operating scheme is shown in Figure 7.3.

7.4. Room compensation

When the sound reproduction system is operating in a non-anechoic host environment, reverberations can drastically affect the quality of the rendered wave field [83].

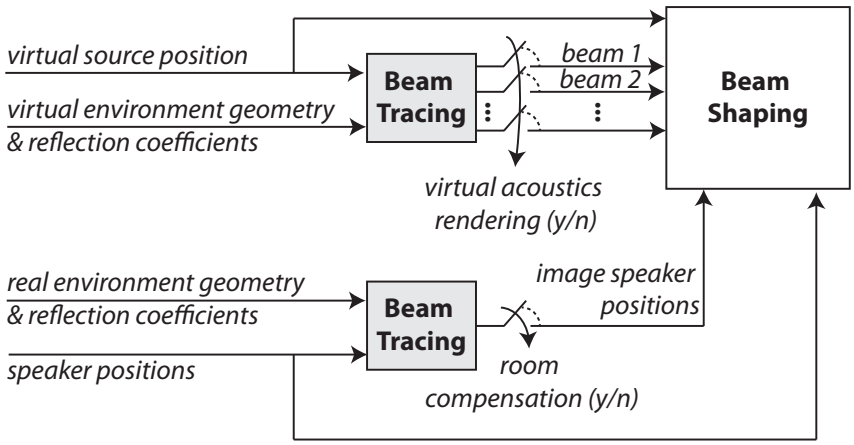


Figure 7.3.: Functional block diagram of the proposed rendering problem.

This degradation is predominantly caused by the early reflections, which affect the desired spatial impression [130]. For this reason, we aim at devising a methodology that focuses on the compensation of the early reflections. Assuming that the two-dimensional map of the real environment is known or estimated using the algorithm presented in Chapter 5, as well as the position of the M loudspeakers $\mathbf{p}_1, \dots \mathbf{p}_M$ and N control points $\mathbf{a}_1, \dots \mathbf{a}_N$, it is possible to predict the impact of the early reflections at the control points. In order to do that we employ the beam tracing engine presented in Chapter 3. More specifically, beam tracing allows us to:

1. determine the set $\{\mathbf{p}'_{m,i}\}_{i=1}^{Q_m}$ of the Q_m image loudspeakers associated with the mth loudspeaker, up to an arbitrary reflection order;

2. evaluate the visibility of the control point $\mathbf{a}_n$ from the image loudspeaker $\mathbf{p}'_{m,i}$; this operation is performed using the path-tracing technique (see Section 3.4.3).

The contribution of the mth loudspeaker to the soundfield at the nth control point can be modeled as

$$\gamma_\omega(\mathbf{a}_n, \mathbf{p}_m) = g_\omega(\mathbf{a}_n|\mathbf{p}_m) + \sum_{i=1}^{Q_m} \beta_{m,i} V(\mathbf{a}_n, \mathbf{p}'_{m,i}) g_\omega(\mathbf{a}_n|\mathbf{p}'_{m,i}) \;, \tag{7.1}$$

where $g_\omega(\mathbf{a}_n|\mathbf{p}_m)$ is the Green's function [131] from $\mathbf{p}_m$ to $\mathbf{a}_n$, $V(\mathbf{a}_n, \mathbf{p}'_{m,i})$ is a binary function that maps the visibility of $\mathbf{a}_n$ from $\mathbf{p}'_{m,i}$. In particular $V(\mathbf{a}_n, \mathbf{p}'_{m,i}) = 1$ if

the image loudspeaker $\mathbf{p}'_{m,i}$ is visible from $\mathbf{a}_n$ and 0 otherwise. The computation of the visibility is swiftly accomplished with beam tracing. The term $\beta_{m,i}$ is the attenuation coefficient associated with the image loudspeaker at $\mathbf{p}'_{m,i}$. The value of $\beta_{m,i}$ depends on the reflective properties of the walls and on the reflection order of the image loudspeaker, and it can be obtained using the algorithm presented in Chapter 6. In equation (7.1) we recognize a first term which corresponds to the free-field propagation (i.e., the Green's function), and a second term including the effect of all the visible reflective paths generated by the mth loudspeaker at the control point $\mathbf{a}_n$.

While rendering a desired beam the room compensation is performed considering, for each loudspeaker and each control point, the propagation (7.1) computed using the beam tracing engine instead of considering just the free-field propagation $g_\omega(\mathbf{a}_n|\mathbf{p}_m)$. The problem of rendering an acoustic beam using a loudspeaker array has been widely investigated in the literature during the past twenty years [132, 133, 134]. The Appendix D summarize the main results from [111]. We remark however that the presented techniques for the room compensation and the rendering of virtual environments are general enough to be employed also with other beam rendering techniques.

7.5. Rendering of the acoustics of virtual environments

Here we formulate the problem of rendering the early reflections coming from walls of a virtual environment. We are interested in rendering early reflections, as they boost the impression of presence to a listener, and therefore are useful for immersive applications. On the other hand late and diffuse reflections provide some acoustic cues related to the dimension of the environment. As shown in [135], late reverberations can be modeled as a common pole, i.e. they are not influenced by the position of the listener. As a consequence, one could enrich the rendering adopting a pre-filter to the dry sound. Nonetheless, this goes beyond the scope of this work.

The output of beam tracing consists of the acoustic beams $b_v, v = 1, \ldots, V$ to be rendered along with the virtual source. The vth beam is characterized by the position of the virtual image source $\mathbf{s}_v$, the orientation θ_v and the aperture ϕ_v. For each beam, we obtain M filters $\hat{h}_m(t), m = 1, \ldots, M$, to be applied to the M loudspeakers in order to render the desired beam. The filters can be computed using the beamshaping algorithm described in Appendix D. In order to render the

soundfield coming from the superposition of all the acoustic beams we apply the superposition principle. More specifically

$$\hat{h}_m(t) = \sum_{v=1}^{V} \hat{h}_m^{(v)}(t - D_v) \ , \tag{7.2}$$

is the filter applied to the mth loudspeaker for the rendering of the V acoustic beams.

Notice that the contribution to the soundfield from the beam b_v must exhibit a delay that is different from that of the beam $b_{v'}, v' \neq v$, due to the different distance of the respective virtual image sources $\mathbf{s}_v$ and $\mathbf{s}_{v'}$. As a result, before applying the superposition principle in (7.2), we need to delay the individual impulse responses $\hat{h}_m^v(t)$ by D_v samples in order to account for the different distances of the virtual image source positions. One could also argue that the superposition principle could be applied on the desired response used for the beamshaping, by summing the desired responses of all the beams. This way, the rendering of the whole soundfield would be simplified. However, a phase alignment of the desired responses of the individual acoustic beams needs to be applied in order to account for the different delays of the virtual image sources, resulting in an improved difficulty with respect to the solution proposed in this section.

Finally, we notice that for common environments, the locations of the virtual image sources surround the listening area. A linear array that covers only one side of the listening room would be able to accurately render only the virtual image sources that are located behind it, resulting in a partial immersivity impression. For this scenario, therefore, circular arrays are required.

7.6. Simulation results

The evaluation of the rendering system and of the room compensation methodology is accomplished through a set of simulation experiments, as commonly adopted in the literature [83, 136, 120, 119, 125]. This is motivated by the fact that a real-world implementation of rendering and/or room compensation techniques would introduce non-idealities (e.g. non linearities and non-omnidirectional loudspeakers), which are hardly measurable and whose effect is not easily distinguishable from the artifacts introduced by the rendering methodology itself. Moreover, as already adopted in [83, 125], acoustic propagation in the real environment is modeled by

an image-source like technique, which in this case is beam tracing. In particular, the room impulse responses of the environment have been simulated up to the 15^{th} order of reflection. The attenuation due to propagation and successive reflections causes higher order beams to be sufficiently attenuated in the listening area, so that their impact can be neglected. On the other hand, in the room compensation stage, we concentrate on compensating reflections up to the 3^{rd} order, as they have the highest impact on the perceived spatial impression [129, 113] and, therefore, would corrupt significantly the desired soundfield if not taken into account.

7.6.1. Evaluation metric

In order to assess the quality of the beam reproduction, we propose to evaluate the mean square error between the desired and rendered wave fields. We define a grid of $Q >> N$ evaluation points that uniformly sample the listening area at positions $\mathbf{q}_i$, $i = 1, \dots Q$, where N is the number of control points used in setting up the beam-shaping algorithm. The metric we consider is the Normalized Mean Square Error (NMSE), defined as [136]

$$E(\omega) = \frac{\sum_{i=1}^{Q} |S(\omega, \mathbf{q}_i) - \hat{S}(\omega, \mathbf{q}_i)|^2}{\sum_{i=1}^{Q} |S(\omega, \mathbf{q}_i)|^2} \,. \tag{7.3}$$

In the following tests, the NMSE associated to the room compensation (E_{rc}) is compared with the NMSE of the beam shaping performed without compensating for the real environment (E_{nc}). Moreover, these results are compared with the NMSE of the beam shaping in the free-field (E_{ff}), i.e. when the real environment is anechoic and no compensation is required.

7.6.2. Rendering of a virtual environment with room compensation

We consider the rendering of a virtual source along with the effect of a virtual environment surrounding it. The simulations are conducted in an L-shaped reverberating real environment. We consider a circular distribution of the loudspeakers, which allows us to render beams originating at any position around the listener. As in Figure 7.4, $M = 32$ loudspeakers are disposed on a circumference with radius $r_a = 1.4$ m. The listening area covers a circle with radius $r_l = 1.1$ m inside the array, and it is regularly sampled with $N = 1000$ control points. For the evaluation, a denser sampling ($Q = 10000$ evaluation points) is considered. We notice that

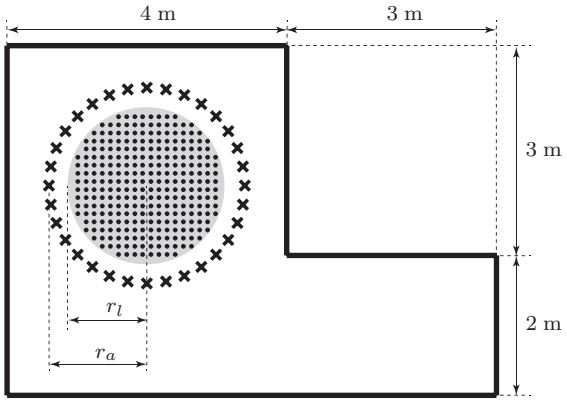


Figure 7.4.: Setup for the evaluation of the rendering engine. The loudspeakers are disposed on a circle with radius r_a and the listening area covers a circle with radius r_l inside the array. The real environment models an L-shaped room.

the circular geometry of the loudspeaker array and of the listening area allow us to assess the maximum alias-free frequency ω_A. In fact, as noticed in [83], an alias-free reproduction at the frequency ω is achieved if $M \geq 2kr_l + 1$, where $k = \omega/c$. For the rendering system used in simulations ($M = 32$ and $r_l = 1.1$ m) we obtain $f_A = \omega_A/2\pi \approx 850$ Hz.

The virtual source and the virtual environment considered for the experiment are shown in Figure 7.5, which depicts also their mutual position with respect to the real environment. It is important to notice that inside the L-shaped real room and especially inside the virtual room that represents a small church, occlusions occur and the visibility of image sources have to be taken into account using the beam tracing method. The reflection coefficient is set to 0.7 which corresponds to the typical value of plaster [105].

The Figure 7.6 shows the NMSE results in the frequency range [100 Hz, 2 kHz]. We focus on this frequency range as in this range the human auditory system is more sensitive to the spatial cues, exhibiting an improved accuracy in sound localization [137]. As expected, looking at the NMSE of the free-field and room-compensated responses, we observe that the wave field is rendered with good accuracy up to the maximum alias-free frequency f_A. Above this frequency value, the quality of rendering decreases because of the spatial aliasing artifacts. On the other hand, the quality of the rendered wave field is poor when no compensation is performed. In fact, the NMSE of the non-compensated response highlights, as usual, the effect of the real environment on the listening area.

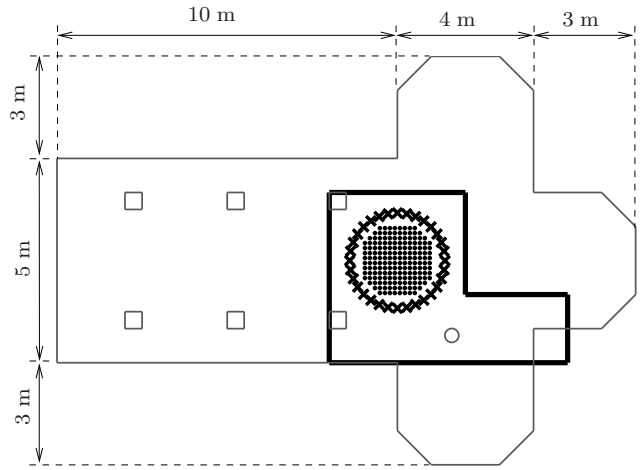


Figure 7.5.: The rendering system operates in an L-shaped real environment and renders a virtual source (depicted as a black circle) along with its effect on a virtual environment that models a small church.

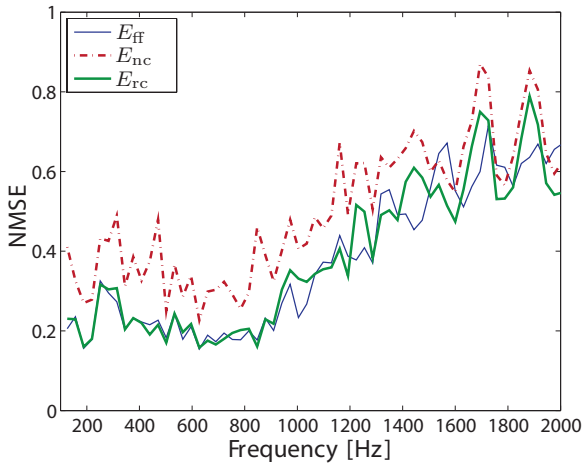


Figure 7.6.: NMSE values of the free-field (E_{ff}), non-compensated (E_{nc}) and room-compensated (E_{rc}) responses of rendering the virtual source and the virtual environment in Figure 7.5.

An example of desired and rendered wave fields is shown in Figure 7.7. More specifically, Figure 7.7 (a) depicts the intensity of the desired wave field, at 850 Hz, relative to the rendering of a virtual source inside the small church (Figure 7.5); Figures 7.7 (b), 7.7 (c) and 7.7 (d) show the free-field, non-compensated and compensated reproduced wave fields, respectively. In order to emphasize the effect of the virtual environment, the direct contribution of the virtual source has been removed from the wave fields. It is worth noticing that free-field and room-compensated

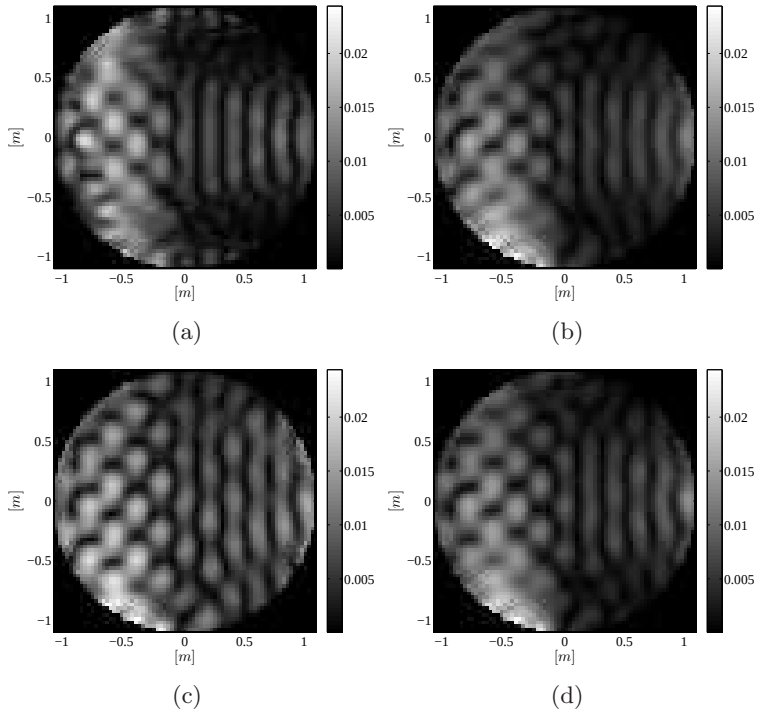


Figure 7.7.: (a) The intensity of the desired wavefield at 850 Hz produced considering the virtual source and the virtual environment in Figure 7.5 is compared with: (b) the intensity of the free-field wave field; (c) the intensity of the non-compensated wave field; (d) the intensity of the room-compensated wave field.

wave fields mostly reproduce the details contained in the desired one. In particular, Figures 7.7 (b) and 7.7 (d) both contain the left-right discontinuity present in 7.7 (a). This discontinuity is due to the two aisles of the virtual church, which partially occlude the reflections coming from the lateral virtual walls (see Figure 7.5). In the non-compensated wave field (Figure 7.7 (c)), such discontinuity is still present, but it is less pronounced. More specifically, most of the details present in the desired wave field (e.g., the "silent" dark spots in the top-right and bottom-right part of the listening area) are lost. This degradation is caused by the strong reflections coming from the left and right walls of the real environment, which, on the other side, are effectively dampened by the room compensation (Figure 7.7 (d)). This behaviour is also confirmed by the NMSE performance, whose values are 0.18, 0.46 and 0.19 for the free-field, non-compensated and compensated wave fields, respectively.

7.6.3. Robustness to geometry and reflection coefficient errors

Now, in order to test the robustness of the system to geometry and reflection coefficient modeling errors, we considered a scenario of a linear array hosted in a 5 m $\times$ 5 m room, as shown in Figure 7.8. The array has an aperture $l = 2\,\mathrm{m}$ and it is composed of $M = 32$ loudspeakers. In the first test the location of the

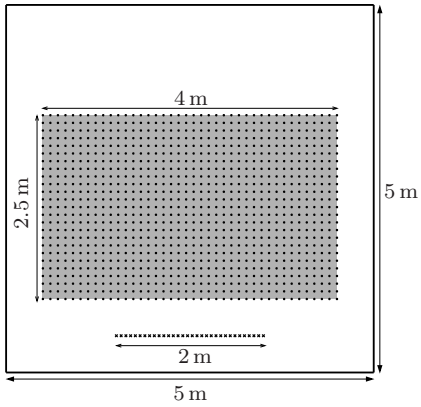


Figure 7.8.: Setup for testing the robustness of the room compensation against variations of the reflection coefficient of the walls and variations of the room size.

walls is assumed to be known. The reflection coefficient is equal for all the walls, but it is known up to a certain precision. More specifically, the actual value of the reflection coefficient is 0.7, but for testing the robustness of the room compensation values from 0.5 to 0.9 have been used in equation(7.1). For the sake of compactly presenting the results, we adopt as metric the average value of $E(\omega)$ over the whole frequency axis. The resulting metric is therefore

$$\overline{E} = \int_0^{2\pi f_{\max}} E(\omega)\, d\omega \,,$$

where $f_{\max}$ is the maximum frequency at which the rendering is performed. In this simulation $f_{\max}$ has been set to 4 kHz. In the following we will refer to $\bar{E}_{\mathrm{ff}}$, $\bar{E}_{\mathrm{nc}}$, and $\bar{E}_{\mathrm{rc}}$ for indicating the average NMSE $\bar{E}$ for the free-field, non-compensated, and room-compensated cases, respectively. For the simulation we considered the reproduction of a beam directed as $\theta = 0^\circ$ (towards the listening area) and aperture $\phi = 10^\circ$, originated from a virtual source located at a distance $d = 5\,\mathrm{m}$ behind the array, positioned such that the center of the beam passes through the array center.

The results are shown in Figure 7.9.

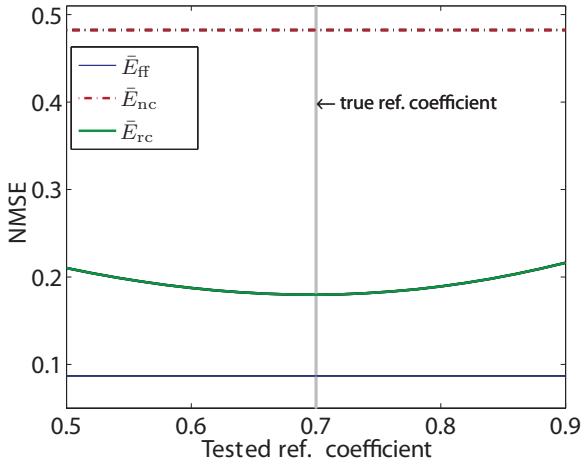


Figure 7.9.: Average NMSE ($\overline{E}$) for values of the reflection coefficient in the range 0.5 $\sim$ 0.9 when the actual reflection coefficient is 0.7.

We conducted a further simulation in which the reflection coefficient used in equation (7.1) has been kept fixed to its nominal value, but the size of the room is known up to a certain precision. More specifically, the length L and the width W spanned in the range 4.8 m $\sim$ 5.2 m. The Figure 7.10 illustrates the variation of $\overline{E}$ as a function of width and length of the room.

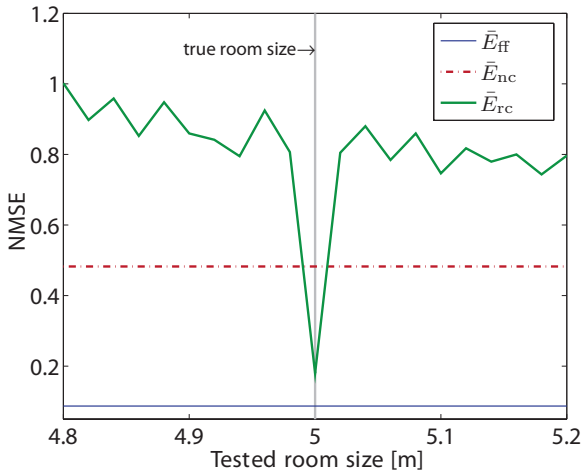


Figure 7.10.: Average NMSE ($\overline{E}$) as a function of the width and length of the room.

From Figure 7.9 we notice that the room compensation exhibits a smooth behavior

over a wide range of reflection coefficients. On the other hand, from Figure 7.10 we notice that even an error of 2 cm in measuring the width and length of the room introduces relevant differences in $\overline{E}$. A similar behavior was outlined in [138]. The reason of this behavior can be found in the fact that when a wrong geometry of the room is in use, the position of the image speakers is wrongly assessed, and this causes a phase shift of the relative contribution in (7.1). Therefore, instead of compensating the reflective path, a constructive interference appears for some frequencies in the listening area. On the other hand, when a wrong value of the reflection coefficient is used, only the amplitude of the reflective paths is wrongly estimated, their delay being correct. This configuration only causes a partial cancelation of the reflective path.

We can conclude that the compensation technique is much more sensible to geometric errors than to the errors in reflection coefficient estimates. This motivates improvements and further research on geometry inference algorithms as a crucial part for advanced audio signal processing algorithms. However, for the geometry inference, other methods, based, for example, on computer vision and telemetry could be used to obtain accuracy well below 1 cm of error at an affordable cost. On the other hand, the reflective properties of walls can be assessed using only the acoustic measurements. In this case, however, the accuracy of the method proposed in the Chapter 6 is suitable for the rendering applications and represents a good alternative to the use of acoustically certified (known reflection coefficient) or acoustically absorbing (working in an low reverberant room) materials, which are often too costly.

7.7. Conclusive remarks

In this Chapter we discussed the problems encountered in applications aimed at rendering the virtual acoustic scenes in reverberant environments. Rendering the acoustics of an virtual environment implies rendering a number of virtual image sources together with their visibility pattern. Working in an reverberant environments means that reverberations will affect the rendered soundfield if not taken into account. In order to compute both the virtual image sources (with their visibility patterns - beams) and the reverberation paths of the hosting environment, we resort to the 2D visibility-based beam tracing algorithm presented in the Chapter 3. The real environment geometry is known or estimated using, for example, the algorithm

presented in the Chapter 5. For the compensation step to work the knowledge of the reflection coefficient is also required. The reflection coefficient of all reflectors in the environment can be estimated using the algorithm described in Chapter 6. For a detailed description, theoretical analysis and exhaustive performance evaluation of beamchaping, compensation and virtual environment rendering techniques see [112]. For a technique that compensate undesired wall reflections while exploiting other early reflections to create a virtual surround system see [53]. For a evaluation metrics that incorporates psychoacoustic-based criteria to analyze the impact of pre-echoes and post-echoes on the perceived quality in soundfield rendering applications see [139].

Chapter 8

Conclusions and future perspectives

In this thesis we proposed a novel representation of plenacoustic information in the parametric domain of the acoustic rays, and we showed that this representation is suitable for approaching challenging problems related to acoustic scene analysis and modeling. In particular, we introduced the ray space representation of plenacoustic data; algorithms for its construction (visibility information) and acquisition (plenacoustic image); as well as a number of applications that make use of plenacoustic data for both modeling and analysis purposes. This last Chapter summarizes the main results of this dissertation and discusses perspectives and future directions of research that can be foreseen to arise from these results. We also discuss possible research problems that need be addressed when using plenacoustic data, the main ideas and challenges that are to be faced in developing solutions for specific tasks.

This thesis addressed a number of topics, from the modeling of acoustic propagation; to the estimation of radiometric and geometric properties of reflective boundaries; to the rendering of virtual soundfields within a reverberant environment; to the construction/acquisition and use of plenacoustic data. All these topics are becoming more and more relevant in modern acoustic systems, as the complexity of the tasks that such systems are facing increase (systems are expected to work in and adapt to acoustically "hostile" environments). In fact, in order to address current and future challenges, the new systems need to employ sophisticated audio signal

processing algorithms and, at the same time, to interact, exchange information and collaborate with each other, by taking advantage of the presence of devices with integrated loudspeaker and microphone arrays in everyday environments (e.g. surround sound systems, "smart TVs", computers, game consoles, smartphones, etc.). The acoustic measurements that are acquired by different systems in different positions in space need to be collected and organized in a parameter space that exhibits a high degree of regularity and can be accessed and used by different signal processing algorithms. Furthermore, in order to guarantee the necessary operational reliability and robustness, and offer a high-quality user experience, the new acoustic systems need to constantly monitor the hosting environment, adapt to it and predict the effects of their actions within it. In other words, the analysis of acoustic scenes and the modeling of acoustic propagation should be carried out in a synergistic fashion, i.e. cooperating between the analysis step, which uses the simulation of the soundfield to improve the results of the analysis, and the modeling step, which uses the extracted information about the environment to simulate the soundfield.

In this thesis the first steps towards a closer integration of spatially distributed systems and of modeling and analysis stages have been performed by defining a suitable parametric space that enables an efficient and accurate modeling of acoustic propagation; efficiently encodes visibility conditions from different points in space; and represents the spatially distributed measurements of the acoustic scene with a high degree of regularity. In particular, this space was found to be the space of oriented lines we called the ray space. On top of this layer we developed both modeling and analysis frameworks, both of which benefit form the high regularity of the ray space representation of the geometric primitives.

The examined applications demonstrated the validity of the proposed approach. First the plenacoustic images were used for geometry inference that, thanks to the fact that the geometric primitives are represented by linear constraints in the ray space, turned out to be a linear problem and the contributions of multiple objects could easily be distinguished (clustered). The plenacosutic images, however, collect and organize all the information gathered from the sensors and not just the geometric ones. This was shown in the second application examined in this dissertation, where an acquired acoustic image was used in combination with the model of acoustic propagation in order to estimate reflection coefficients of all reflective surfaces found within the environment. Finally, the third application acted as a framework for previous ones; it used the geometric information and the reflection coefficients as

inputs of the modeling engine in an advanced rendering scenario.

The generality and the potential of the main ideas behind tools developed in this dissertation go far beyond the examined applications. In this preliminary stage the obtained plenacoustic images have limited resolution given by the limited number of sensors and a considerable distortion generated by the measurement method (i.e. the beamforming filter does not completely suppress the signals coming from directions different from the desired one resulting in an error in the estimation of the sound field carried by the given acoustic ray). Nonetheless, the plenacoustic images they capture both geometric and radiometric properties of the environment and exhibit a number of interesting features and, therefore, have numerous potential applications. The appeal of the plenacoustic representation grows even further if we consider how rapidly new devices with integrated microphone arrays are evolving and how quickly they are penetrating the consumer market. At the same time the literature is introducing relevant algorithmic improvements for audio data processing (e.g. superdirective beamforming techniques), which could be fruitfully integrated in the vision proposed in this thesis and could serve as technological basis for their implementation. However, as we explore new applications, new challenges arise. In what follows, we briefly discuss some of these new ideas and challenges that lie ahead.

- **Self-calibration** – Given a number of spatially distributed microphone arrays, in order for them to work together their mutual positions have to be known a priori. If not so, the arrays have to estimate their mutual positions (self-calibrate themselves). This calibration can be achieved by matching of plenacoustic images acquired by different arrays, i.e. by finding the translation and rotation in the ray space so that the features that are present in different images will closely match. This way the calibration can be performed without the need to estimate the acoustic scene (e.g. localize sources) prior to the calibration step. Calibration can be achieved independently of acoustic primitives that compose the acoustic scene as long as the scene exhibits a sufficient "plenacoustic texture" to enable the matching of soundfield images acquired with different arrays. Notice, however, that in this case the plenacoustic images cannot be represented in a reduced ray space as this requires the intersection with a prescribed plane that cannot be the same across all arrays (mutual positions are unknown). The plenacoustic images, therefore,

need to be represented in the full (projective) ray space.

- **Wavefield extrapolation** – Given a series of plenacoustic images captured by a linear microphone array at different time intervals ("plenacoustic video") this information can be used for the extrapolation of the wavefield at a generic position behind the array. In fact the "virtual receiver" is represented as a line in the image and, therefore, it should capture the acoustic rays distributed along this line. However, for purposes of room geometry estimation, the phase information was discarded considering the output power of the beamformer for the construction of the plenacoustic images. On the other hand, for purposes of wavefield extrapolation the phase information has to be re-introduced by considering the phase of output signals of the beamforming technique. Furthermore, the relation between this "phase image" and the position of the virtual receiver has to be examined.

- **"Plenacoustic" reflection coefficient estimation** – In this thesis we estimated the reflection coefficients using measurements from a single point in space. Using a number of spatially distributed acoustic images of this sort (plenacoustic image) we could estimate the reflection coefficients without using the hypothesis that they remain constant along the same wall. In fact, the walls of real environments have doors, windows, posters, blackboards, etc. and therefore they do not exhibit constant reflective properties.

- **Source characterization** – Observing the radiation emitted by an acoustic source from different positions in space, i.e. along the corresponding line in the plenacoustic image, the radiation pattern of the source could be estimated.

- **Source separation** – Sources are mapped onto lines in the plenacoustic image. By observing the image we can get an idea of which directions certain sources are most clearly received and which directions they are expected to mix the most (intersections of lines). With this information at hand (and, possibly, the estimated radiation patterns) it would be possible to build a "matched filter" (sort of a plenacoustic mold) for a given source, which extracts the signal from the directions that offer the most favorable signal-to-interference ratio. As in the case of wavefield extrapolation, the phase information should be considered as well.

- **"Plenacoustic" rendering** – In this dissertation we considered a model-based rendering scenario. However, given the plenacoustic image of a certain environment acquired by a microphone array, the acoustics of this environment could be reproduced elsewhere using a loudspeaker array, i.e. substituting the microphone array with a loudspeaker array and reversing the plenacoustic image acquisition process into a plenacoustic reproduction process.

In conclusion, the applications proposed and examined in this dissertation prove the potential of using the ray space representation of spatially distributed data for both modeling of sound propagation in enclosures and analysis of acoustic scenes. Furthermore, a number of other applications could benefit from the use of the tools that are developed in this thesis and the first works are currently being carried out in these directions. The main ideas that inspired this work, namely environment awareness; collaboration between different spatially distributed systems; and alternating analysis and modeling steps; are becoming increasingly important topics in the audio signal processing. They represent important assets that could be exploited in order to deal with the growing complexity of task that acoustic systems are asked to perform. We believe that the approach presented in this work can give a significant contribution to research in the above fields and inspire novel and innovative solutions.

Appendix A

Wavefield and rays: the eikonal equation

In the first part of this Appendix we derive the wave equation that governs the propagation of acoustic waves. To do so we use two fundamental principles of mechanics: the principle of inertia and Hooke's law. In the second part of the Appendix we introduce a particular high-frequency approximation of the wave equation called eikonal equation on which all geometric methods for the solution of the wave equation are based.

A.1. Derivation of the acoustic wave equation

The wave equation is a second-order partial differential equation that governs the propagation of waves through a medium. In its most general form it describes the propagation of all types of waves (e.g. sound waves, light waves, seismic waves and water waves). Therefore its study is important in fields such as acoustics; electromagnetics; seismic; and fluid dynamics. Historically, the theory behind the acoustic wave equation [140], [141] was developed first. It was studied, among others, by Sir Isaac Newton; Pierre-Simon Laplace; Daniel Bernoulli; René Descartes; Christiaan Huygens; Leonhard Euler; Jean le Rond d'Alembert and Joseph-Louis Lagrange.

The acoustic field is described by acoustic pressure and particle velocity. Pressure can be thought as sum of two terms, static pressure p_0 (usually controlled by gravity)

and variations due to the presence of the wave p, i.e.

$$p_t = p_0 + p,$$

and the same for density, sum of static term ϱ_0 and variations caused by the wave ϱ, i.e.

$$\varrho_t = \varrho_0 + \varrho.$$

We suppose that perturbations are due to acoustic wave only. In this condition the wave equation is derived from two fundamental principles of mechanics: principle of inertia (Newton's first law of motion) and Hooke's law of elasticity. The principle of inertia makes particles velocity change with changes of pressure, i.e variations of pressure in space generate variations of velocity in time. Hooke's law yields changes of pressure as a result of changes of velocity. As a consequence these two mechanisms generate a wave.

A.1.1. Principle of inertia

Let us suppose that a force $\Delta\mathbf{F}$ is applied to a fluid of volume ΔV and mass Δm. We can write the principle of inertia as

$$\Delta\mathbf{F} = \Delta m \frac{\partial \mathbf{v}}{\partial t}. \tag{A.1}$$

A force can be expressed as a pressure

$$\Delta F_x = \frac{\partial p}{\partial x}\Delta x \Delta S_x = -\frac{\partial p}{\partial x}\Delta V,$$

that, extended to all directions, becomes

$$\Delta\mathbf{F} = -\nabla p \Delta V,$$

where ∇p denotes the gradient of p. Inserting this expression of force in (A.1) we get

$$-\nabla p = \frac{\Delta m}{\Delta V}\frac{\partial \mathbf{v}}{\partial t},$$

or

$$-\nabla p = \varrho_0 \frac{\partial \mathbf{v}}{\partial t}. \tag{A.2}$$

The equation (A.2) links, according to the principle of inertia, variation of pressure in space with variation of velocity in time. The two forces are in equilibrium.

A.1.2. Hooke's law

First we need to link the density and volume using the principle of mass conservation

$$\Delta m = \varrho \Delta V = constant. \tag{A.3}$$

Perturbing the mass, expressed as product of density and volume, we get

$$\varrho \Delta V = (\varrho + d\varrho)(\Delta V + dV).$$

Expanding and ignoring the infinitesimals of second order we have

$$\frac{d\varrho}{\varrho} = -\frac{dV}{\Delta V}. \tag{A.4}$$

Next we need to link density variations with pressure variations using adiabatic transformations. The mathematical equation for an ideal fluid undergoing a reversible (i.e. no entropy generation) adiabatic process is

$$p\Delta V^{C_p/C_V} = constant,$$

or, using (A.3)

$$p\varrho^{-C_p/C_V} = constant,$$

where C_p and C_V are specific heat for constant pressure and specific heat for constant volume, respectively. Their ratio, k, is called adiabatic index. Elevating to $-1/k$ we get

$$p^{-1/k}\varrho = constant,$$

perturbing

$$p^{-1/k}\varrho = (p + dp)^{-1/k}(\varrho + d\varrho) = p^{-1/k}(1 + \frac{dp}{p})^{-1/k}\varrho(1 + \frac{d\varrho}{\varrho}),$$

and linearising

$$p^{-1/k}\varrho = p^{-1/k}(1 - \frac{1}{k}\frac{dp}{p} + ...)\varrho(1 + \frac{d\varrho}{\varrho}),$$

we finally obtain

$$\frac{d\varrho}{\varrho} = \frac{1}{k}\frac{dp}{p}.$$

Using this expression in (A.4) we have

$$\frac{1}{k}\frac{dp}{p} = -\frac{dV}{\Delta V},$$

$$dp = -kp\frac{dV}{\Delta V}.$$

The term kp is known as bulk modulus K. We obtain

$$dp = -K\frac{dV}{\Delta V}. \tag{A.5}$$

From (A.5) we see that as a result of a volume compression $-dV$ we obtain a increase of pressure dp and vice versa.

The volume perturbation can be expressed as

$$dV = dx\Delta y\Delta z + dy\Delta x\Delta z + dz\Delta x\Delta y,$$

and

$$\frac{dV}{\Delta V} = \frac{dx}{\Delta x} + \frac{dy}{\Delta y} + \frac{dz}{\Delta z}.$$

The stretching dx can be written as

$$dx = (v_x dt)_{x+\Delta x} - (v_x dt)_x = \frac{\partial v_x dt}{\partial x},$$

where v_x, v_y and v_z are components of the velocity v in directions x, y and z. Writing the same for other directions we obtain

$$\frac{dV}{\Delta V} = (\nabla \cdot \mathbf{v})dt,$$

where $\nabla \cdot \mathbf{v}$ is the divergence of $\mathbf{v}$. Using this expression in (A.5) we write

$$dp = -K(\nabla \cdot \mathbf{v})dt,$$

$$-\nabla \cdot \mathbf{v} = \frac{1}{K}\frac{\partial p}{\partial t}. \tag{A.6}$$

The equation (A.6) represents the Hooke's law for waves in fluids. It states that

the variation of velocity generate the variation of pressure.

A.1.3. Acoustic wave equation

From (A.2) and (A.6) we can derive the acoustic wave equation [140], [141]. They are analogous to Maxwell's equations used to derive the wave equation for electromagnetic waves. Applying the divergence operator to both members of (A.2) we obtain

$$\nabla \cdot (-\frac{1}{\varrho_0}\nabla p) = \nabla \cdot (\frac{\partial \mathbf{v}}{\partial t}) = \frac{\partial(\nabla \cdot \mathbf{v})}{\partial t},$$

that, after using (A.6), becomes

$$\nabla \cdot (-\frac{1}{\varrho_0}\nabla p) = \frac{\partial}{\partial t}(-\frac{1}{K}\frac{\partial p}{\partial t}),$$

and, after multiplying by the density, we get

$$\varrho_0 \ \nabla \cdot (\frac{1}{\varrho_0}\nabla p) - \frac{1}{c^2}\frac{\partial^2 p}{\partial^2 t} = 0, \tag{A.7}$$

where c is the propagation velocity given by

$$c = \sqrt{\frac{K}{\varrho_0}}.$$

Remembering that

$$\nabla \cdot (\phi \mathbf{A}) = \mathbf{A} \cdot (\nabla \phi) + \phi(\nabla \cdot \mathbf{A}),$$

$$\nabla \frac{1}{\phi} = -\frac{1}{\phi^2}\nabla \phi,$$

and

$$\nabla \ln \phi = \frac{1}{\phi}\nabla \phi,$$

we can write

$$\nabla \frac{1}{\phi} = -\frac{1}{\phi}\nabla \ln \phi.$$

Equation (A.7) becomes

$$\varrho_0(-\nabla p \cdot (\nabla \ln \varrho_0)\frac{1}{\varrho_0} + \frac{1}{\varrho_0}\nabla \cdot (\nabla p)) - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0,$$

and simplifying

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - (\nabla \ln \varrho_0) \cdot \nabla p = 0.$$

The last term represents the variations of static density from point to point. It causes distortions of spacial variations of pressure. It can be proven that if these variations of density are much smaller than 2π of wavelength (this assumption is more likely to be true on high frequencies), i.e. if

$$|\nabla \ln \varrho_0|\lambda << 2\pi,$$

we can ignore this term and, finally, write the acoustic wave equation as

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0. \tag{A.8}$$

This equation governs the propagation of acoustic waves through liquid mediums (air, water , etc.). In uniform and homogeneous environments the solution of this wave equation is simple. However, in complex environments in which are present a number of mediums with different acoustic properties the solution of the wave equation is governed by boundary conditions becoming a boundary value problem which solution is usually complex. In order to solve such a boundary value problem a number of approximations are made.

A.2. Eikonal equation

Let us explain more in detail the high frequency approximation of the wave equation, its advantages and disadvantages. In its most general form the Fourier transform of the solution of acoustic wave equation (A.8) is

$$P(\mathbf{x}, \omega) = S(\omega)A(\mathbf{x}, \omega)\exp(j\omega T(\mathbf{x})), \tag{A.9}$$

where $S(\omega)$ is the Fourier transform of the source's signal. The amplitude $A(\mathbf{x}, \omega)$ depends on both position and frequency. The phase $T(\mathbf{x})$ depends only on position. This is, clearly, an approximation. If we use $T(\mathbf{x}, \omega)$ there would be no approximation. However, the use of $T(\mathbf{x})$ independently from frequency is fundamental as this condition implies the existence of the wave-front. In order to verify this condition the spectral dispersion of the signal needs to be low. The phase function $T(\mathbf{x})$ is

called eikonal.

We can approximate $A(\mathbf{x},\omega)$, separating the dependency from position and frequency, as

$$A(\mathbf{x},\omega) = A_0(\mathbf{x}) + \frac{A_1(\mathbf{x})}{j\omega} + \frac{A_2(\mathbf{x})}{(j\omega)^2} + ...,$$

and take only the first term $A_0(\mathbf{x})$ (this approximation is true only on high frequencies in which other terms, inversely proportional to frequency, can be ignored). With these approximations the (A.9) becomes

$$P(\mathbf{x},\omega) = S(\omega)A_0(\mathbf{x})\exp(j\omega T(\mathbf{x})), \tag{A.10}$$

that, in the space-time (x,t) domain, represents

$$p(\mathbf{x},t) = A_0(\mathbf{x})s(t - T(\mathbf{x})),$$

i.e. $s(t)$ delayed by the travel time T, propagated without distortions.

The wave equation (A.8) can be written as Helmholtz's equation [140], [142], [2]

$$\nabla^2 P(\mathbf{x},\omega) + \frac{\omega^2}{c^2}P(\mathbf{x},\omega) = 0. \tag{A.11}$$

Using (A.10) we can write

$$\nabla P = S\nabla A_0 e^{j\omega T} + SAj\omega\nabla T e^{j\omega T},$$

$$\begin{aligned}\nabla^2 P &= S\nabla^2 A_0 e^{j\omega T} + S\nabla A_0 j\omega\nabla T e^{j\omega T} + j\omega S\nabla A_0\nabla T e^{j\omega T} + \\ &\quad + j\omega S A_0\nabla^2 T e^{j\omega T} - \omega^2 S A_0(\nabla T)^2 e^{j\omega T},\end{aligned}$$

and thus, after eliminating common terms, we obtain

$$\nabla^2 P + \frac{\omega^2}{c^2}P = S\nabla^2 A_0 + 2j\omega S\nabla A_0\nabla T + j\omega S A_0\nabla^2 T - \omega^2 S A_0(\nabla T)^2 + \frac{\omega^2}{c^2}SA_0 = 0.$$

After ordering the terms and eliminating S we obtain three equations

$$(\nabla T(\mathbf{x}))^2 - \frac{1}{c^2(\mathbf{x})} = 0, \tag{A.12}$$

$$2A_0(\mathbf{x})\cdot\nabla A_0(\mathbf{x})\cdot\nabla T(\mathbf{x}) + A_0^2(\mathbf{x})\cdot\nabla^2 T(\mathbf{x}) = 0, \tag{A.13}$$

and

$$\nabla^2 A_0(\mathbf{x}) = 0, \tag{A.14}$$

where the equation (A.12) depends on ω^2, (A.13) depends on ω and (A.14) is constant with frequency. The linear and constant terms can be ignored for very high frequencies. As a consequence, only equation (A.12) governs the propagation. It is called eikonal equation and in this particular approximation represents the Helmholtz's equation. However, while in (A.11) the operator $[\nabla^2 + \frac{\omega^2}{c^2}]$ operates on all directions of $p(\mathbf{x})$, $\nabla T(\mathbf{x})$ operates only in the direction of maximum variation of $T(\mathbf{x})$, i.e. the direction of propagation. Thus the ray, i.e. the solution of the eikonal equation, is insensitive to eventual medium variations orthogonal to the direction of propagation, which actually influences the propagation. As a consequence, the solution derived from (A.12) is more accurate on higher frequencies.

Eikonal equation is the basis of the ray theory [141], [143] widely used in seismic and geophysics applications [144], [145]. It is also the base of the geometrical (ray) acoustics [2] employed in the geometric methods for the modeling of the acoustic propagation in complex environments [11], [12], [13], [14].

Appendix B

Geometric approaches to the solution of the wave equation

THE geometric methods for modeling of propagation in enclosures were first developed in the geometrical (ray) optics and then applied in acoustics. Often a variation of the ray tracing technique, these methods are conceptually simple but not entirely physically accurate as they ignore some wave properties. Here we review the most famous ones.

B.1. From computer graphics to acoustic modeling

Geometric methods for the solution of the wave equation follow high frequency approximations based on Eikonal equation (Appendix A). Initially geometric methods were developed as global illumination algorithms in computer graphics and only successively applied for acoustic modeling. However, sound and light have some important differences. For example, sound has longer wave-lengths than light. As a consequence, diffraction is significant; specular reflections dominate diffuse reflections and occlusions by small objects have little effect. Furthermore, sound waves are coherent and thus modelling phase is important. Sound also travels more slowly that light and making reverberations perceived over time. Geometric methods consider the propagation of sound through the air in straight lines, rays, avoiding the wave nature of the sound. The phenomena due to the wave nature (e.g. diffrac-

tion and diffusion) are usually ignored in the ray approach. This approximations are good enough only on high frequencies and thus, in order to make this methods valid also on lower frequencies, different modifications need to be made. Nonetheless the geometric methods are the most widespread techniques for the modeling of early acoustic reflections in complex environments. The alternative approach is based on an approximate solution of the wave equation on a finite grid. Typical examples of this class of methods are based on the solution of the Green's or Helmholtz-Kirchoff's equation through finite (FEM) and boundary element (BEM) methods [16], [17], [18]. Though these methods provide an accurate solution, the storage space and computation time increase dramatically with frequency due to the sampling step requirements (it has to be much smaller that the size of the wavelength), making these algorithms unsuitable for real-time operation except for a very limited range of frequencies.

B.2. Ray tracing

The first geometric methods applied in acoustic rendering have been borrowed from computer graphics. The first one was based on ray tracing [11], [20], [21], [22]. In Figure B.1 is shown an example of ray tracing. A finite number of rays are cast from the source in all directions. They interact with the environment, reflect and attenuate. Rays that, during the propagation, come in the proximity of the receiver determine the room impulse response in the receiver's position. This impulse response can successively be used for auralisation by convolving source's sound signal with the impulse response at receiver position.

The main advantage of this method is its simplicity. It is very simple to implement and can easily model all types of reflections (specular reflections, diffusive reflections and diffraction) as well as all types of surfaces (including curved ones). On the other hand it turns to be quite inefficient. Observing the example in Figure B.1 (a) we notice that rays spread out over distance and only a small number of rays cast from a source pass in proximity of the receiver. Furthermore, due to the discrete sampling of rays, under-sampling errors in the estimated room responses are possible. Observing the example in Figure B.1 (b) we notice that there are no direct paths between source and receiver and there is only one first order reflective path (dashed line) missed by all traced rays (solid lines). As a consequence if the number of rays is not high enough (generating a huge quantity of data for computing and decreasing

performance), some important propagation paths could be missed by all sample rays.

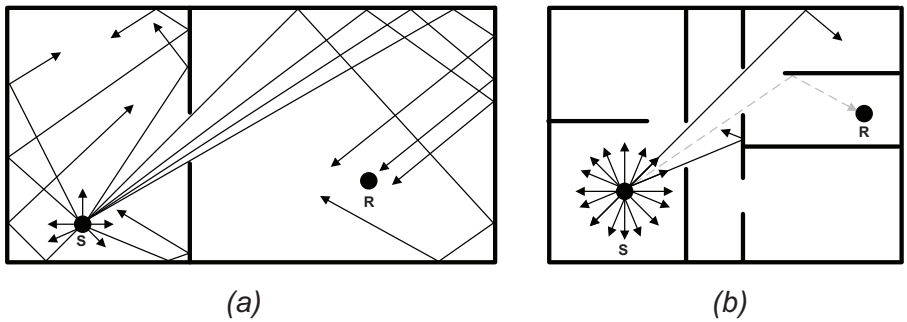


Figure B.1.: Ray tracing: rays are cast from the source in all directions; (a) only a small number reach the receiver; (b) some important propagation paths could be missed by all sample rays.

B.3. Radiosity

Radiosity algorithm [146], [147], [148], [149], [150], was first developed in the field of heat transfer, used as global illumination algorithm for computer graphics rendering and then applied in acoustics. The key assumption is the ideal diffuse (Lambertian) reflection. Radiosity computes rays generated by the source that, after a certain number of diffuse reflections, reach the receiver. It is essentially an application of the finite element method. The environment surfaces are subdivided into small surface elements - patches. The radiosity B_i is the energy leaving the patch surface i. It is a sum of emitted, E_i, and reflected energy. The reflected energy depends on reflectivity of the patch ρ and the incident energy arriving at the patch, i.e. the radiosity of all other patches B_j multiplied by a corresponding form-factor F_{ij} (see Figure B.2)

$$B_i = E_i + \rho \sum B_j F_{ij}. \tag{B.1}$$

Form-factors F_{ij} depend on the mutual visibility between patches (e.g. patches far from each other or partially occluded patches have reduced form-factors while the form-factor of completely occluded patches is zero). Writing (B.1) for all patches gives a system of linear equations whose solution is the radiosity for all patches. Though different methods can be used to solve this equations, the solution depends

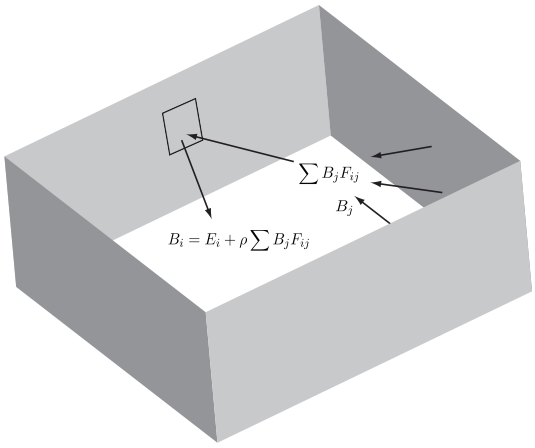


Figure B.2.: Radiosity: the energy leaving the patch surface is a sum of emitted and reflected energy; reflected energy depends on reflectivity and incident energy; incident energy is a fraction of the energy leaving other patches.

on frequency (i.e. the reflectivity terms depend on frequency) and thus must be solved for different frequencies of interest. This calculations are independent of the receiver position, and thus the most of the computational effort is done before the position of the receiver is specified, making radiosity "view-independent" unlike ray tracing. In this way the effectiveness of the algorithm is improved. Although radiosity has been successfully applied in computer graphics it is less popular in acoustics for a number of reasons. Even if some extensions have been made in a effort to include specular reflections, radiosity is based on the assumption of ideal diffuse reflection but in acoustics specular reflections are far more common than diffuse reflections. Furthermore, as sound waves are coherent modelling phase is important and the time dependence must be introduced. All this makes radiosity more difficult to implement and computational demanding in acoustics.

B.4. Image tource

The image source method [12], [23], [24], improves the ray tracing method. The acoustic source is interactively mirrored to the other side of all reflectors that constitute the environment (see Figure B.3 (a)). The mirrored sources are called image sources. Image source determines acoustic field in an opposite portion of space (with respect to the reflector used as mirror). Once the position of the receiver is known, with a test we find acoustic paths linking image sources with the receiver.

Doing so, unlike ray tracing, all specular paths up to the desired reflection order are found. It is extremely efficient for simple, rectangular rooms, like the one in Figure B.3 (a). However, in a densely occluded environments it turns to be quite demanding from computational point of view. In presence of more complicated geometries the problem is that the source is interactively mirrored to the other side of all reflectors generating approximately $(n-1)^r$ image sources (n being the number of reflectors and r the reflection order) many of which are actually invisible, either from a reflected source (i.e. do not generate any acoustic field) or receiver (see some examples in Figure B.3(b), (c) and (d)). As a consequence, time and memory is consumed in futile calculations (on one hand, the number of image sources grow exponentially with the increase of the reflection order, but on the other hand, in complex environments only a small portion of them is actually visible making most of the effort vain). Furthermore a demanding test is required to assess the visibility of all traced image sources, i.e. all image sources should be tested for occlusion by all potential occluders, which combined with high number of reflectors and image sources makes image source method computationally demanding in complex environments. Another drawback of the image source method is that it models only specular reflections.

B.5. Beam tracing

Among the geometric solutions particularly efficient is that of beam tracing [25], [26], [27], [28], [13]. This method was originally conceived by Hanrahan and Heckbert [26] for applications of image rendering, and was later extended by Funkhouser et al. [14] to the problem of audio rendering. The beam tracing method takes advantage of spatial coherence, i.e. it groups rays into compact bundles called beams. All reflectors are assumed to be piecewise planar, and all the rays originated from a source, which hit the same planar region of the reflector are bundled up into beams (see Figure B.4 (a)). The interactive mirroring of the source is done only with respect to the reflectors that a beam encounters during the propagation. In this way the time for the construction of the beam data structure is reduced (compared to image source method in which we construct a similar structure mirroring all image sources) and a minimal number of image sources is computed.

Observing Figure B.4 (a) and (c), we see how every time a beam encounters a reflector, the portion of beam that illuminates that reflector splits into a set of sub-

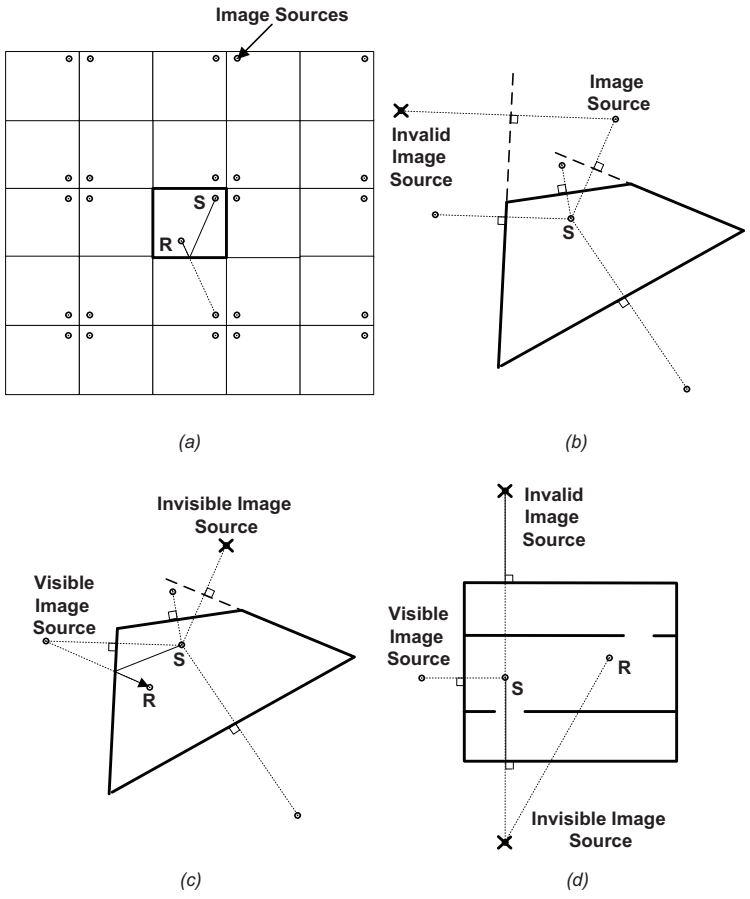


Figure B.3.: Image source method: (a) image sources arranged in grid pattern; (b) not all reflections produce valid image sources; (c) not all image sources obtained mirroring the source are actually visible from the receiver; (d) valid/invalid and visible/invisible image sources.

beams, each corresponding to a different planar portion of the encountered reflector. As they bounce around in the environment, beams keep branching out and attenuating until they die out. The beam-tracing method organizes and encodes this beam splitting/branching process into a specialized data structure called beam tree, shown in Figure B.4 (b) and (d). The construction of the beam tree is based on an iterative visibility evaluation process, usually based on spatial subdivision. It is quite apparent from the above discussion that beams can be cast only with the knowledge of the source location and the environment's geometry. When the receiver is specified as well, the paths linking source and receiver can be determined. Every time a beam falls onto the receiver, we have a path linking source and receiver. The path can be readily determined using the sole information stored in the beam tree. In fact,

given the location of the receiver, we can immediately determine which beams illuminate it, just through a look-up of the beam tree data structure. Thus, the beam tracing approach enables a real-time rendering of sounds in complex environments even when receivers are moving.

Observe the example of beam tracing shown in Figure B.4(a). We can notice that the beam $\mathbf{b}_3$ illuminates only partially the reflector $\mathbf{r}_3$. If we use the image source method instead of beam tracing, we compute only the image source obtained mirroring the source with respect to $\mathbf{r}_3$, and eliminate the paths going from the non illuminated region of $\mathbf{r}_3$ only during the path tracing phase. Using beam tracing we eliminate occluded paths during the construction of the beam tree. Image sources that are never visible are automatically eliminated in this process and thus we avoid successive tests for evaluation of effective visibility of computed image sources necessary in the image source method. In Figure B.4(b) the corresponding beam tree is visualized. Figures B.4(c) and (d) show the splitting process of the beam $\mathbf{b}_3$.

The main disadvantage of the beam tracing method is the fact that if the source moves the beam tree has to be recalculated, which is generally a computationally demanding task. Beam tracing method turns out to be also quite complex to implement. Furthermore, the beam tracing was originally conceived for the modelling of specular reflections only. Early reflections are known to carry some information on the geometry of the surrounding space and on the spatial positioning of acoustic sources. It is in the initial phase of reverberation, in fact, that we receive the echoes associated to the first wall reflections. Other propagation phenomena, such as diffusion, transmission and diffraction tend to enrich the sense of presence in "virtual walkthrough" scenarios, especially in densely occluded environments. Therefore, some extensions of beam tracing method were proposed to account for other propagation phenomena. Funkhouser et al. [29], for example, account for diffusion and diffraction through a Bidirectional Beam Tracing process. When the two beam trees that originate from the receiver and the source intersect on specific geometric primitives such as edges and reflectors, propagation phenomena such as diffusion and diffraction could take place. A different approach was proposed by Tsingos et al. [151], who proposed to use the Uniform Theory of Diffraction (UTD) [152], derived from the Geometric Theory of Diffraction (GTD) [153], by building secondary beam trees originated from the diffractive edges. This approach is quite efficient, as the tracing of the diffractive beam trees can be based on the sole geometric configuration of reflectors. Once source and receiver locations are given, in fact, a simple

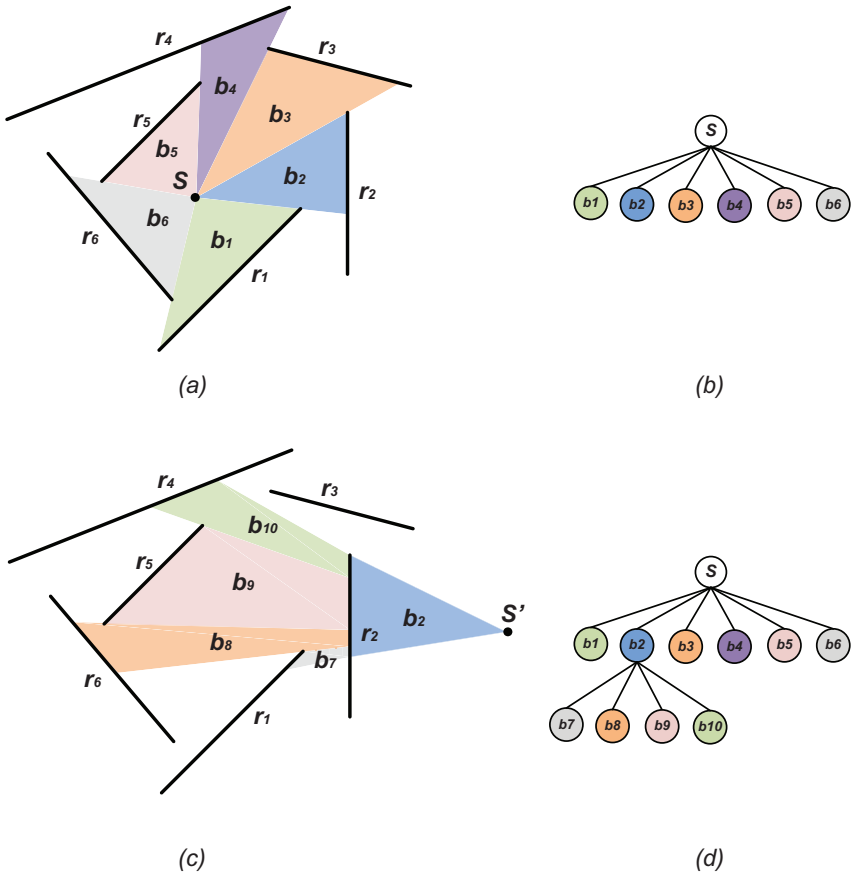


Figure B.4.: Beam tracing: (a) initial splitting of source's acoustic field into beams; (b) corresponding beam tree; (c) the splitting process of the beam $\mathbf{b}_3$; (d) construction of the beam tree.

test on the diffractive beam trees determines the diffractive paths.

Appendix C

Beamforming techniques

In this Appendix we consider both data-independent and statistically optimum beamformers, namely the delay-and-sum (DAS) and the Minimum Variance Distortionless Response (MVDR) beamformers [91], [76], [77]. For the case of wideband sources we describe the wideband MVDR beamformer [75]. Notice that the subspace methods such as the Multiple Signal Classification (MUSIC) and Min-Norm methods [154, 155, 156, 157] can not be applied if the signals are coherent (for more details see [76] 6.4.3 and 6.4.4), which is the case when we consider acoustic reflections from the boundaries of the environment.

Given an wideband source and an M-element microphone array, the signals acquired by the sensors are $x_j(t), j = 1, \ldots, M$. The first step is to operate a filtering in order to obtain $x_j(t, f_k), k = 1, \ldots, K$, where f_k is the kth central frequency of the kth sub-band. Such signals are stacked into the vector $\mathbf{x}(t, f_k) = [x_1(t, f_k), ..., x_M(t, f_k)]^T$.

We then compute the autocorrelation matrix

$$\hat{\mathbf{R}}_k = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}(t, f_k)\mathbf{x}(t, f_k)^H.$$

The pseudospectrum of the kth sub-band is

$$P_k(\theta) = \mathbf{h}^H(\theta, f_k)\hat{\mathbf{R}}_k\mathbf{h}(\theta, f_k), \tag{C.1}$$

where $\mathbf{h}(\theta, f_k)$ is the spatial filter of the given beamforming technique for frequency f_k and direction θ.

C.1. Delay-and-sum beamformer

The pseudospectrum of a delay-and-sum (DAS) beamformer is obtained by substituting $\mathbf{h}(\theta, f_k) = \mathbf{a}(\theta, f_k)$ into (C.1), which yields

$$P_k(\theta) = \mathbf{a}^H(\theta, f_k)\hat{\mathbf{R}}_k\mathbf{a}(\theta, f_k),$$

where the array steering (propagation) vector $\mathbf{a}(\theta, f_k)$ is given by

$$\begin{aligned} \mathbf{a}(\theta) &= [a_1(\theta, f_k), a_2(\theta, f_k), ..., a_M(\theta, f_k)]^T \\ &= \frac{1}{\sqrt{M}}\left[1, e^{j2\pi f_s(\theta)}, ..., e^{j(M-1)2\pi f_s(\theta)}\right]^T, \end{aligned} \tag{C.2}$$

and $f_s(\theta)$ denotes the spatial frequency that is dependent on the angle of arrival θ, temporal frequency f_k, sound speed c, and array geometry.

The main advantage of the DAS beamformer is that, as a data-independent beamformer design that maximizes SNR for spatially white noise, it is robust against microphone self-noise and array mismatches, i.e., the deviations in the positioning of the microphones. On the other hand, it provides limited spatial selectivity at low frequencies.

C.2. MVDR beamformer

In order to improve spatial selectivity for low and middle frequency ranges, superdirective beamformers can be used. MVDR (called also Capon method) is such a statistically optimum design, where the output variance (or power) is minimized subject to a distortionless constraint on the response of the beamformer in the look direction [158]. The closed-form solution of the array weight vector is given by

$$\mathbf{h}(\theta, f_k) = \frac{\hat{\mathbf{R}}_k^{-1}\mathbf{a}(\theta, f_k)}{\mathbf{a}^H(\theta, f_k)\hat{\mathbf{R}}_k^{-1}\mathbf{a}(\theta, f_k)}. \tag{C.3}$$

Substituting (C.3) into (C.1), the measured MVDR pseudospectrum for the kth

sub-band is

$$P_k(\theta) = \frac{1}{\mathbf{a}^H(\theta, f_k)\hat{\mathbf{R}}_k^{-1}\mathbf{a}(\theta, f_k)}. \tag{C.4}$$

The wideband MVDR pseudospectrum [75] is given by

$$P(\theta) = \prod_{k=1}^{K} P_k(\theta).$$

C.3. Focusing matrices and frequency smoothing

Superdirective beamformers are highly sensitive to the microphone self-noise and array errors, it is therefore desired to control the robustness of the design, which can for instance be achieved using the diagonal loading with the frequency-dependent loading factor obtained from iterative design schemes [159]. In addition, for extraction of room reflection signals that have low energy and are strongly correlated with the direct-path signal, focusing matrices and frequency smoothing techniques can be used to alleviate the problem of ill-conditioning of the microphone autocorrelation matrix $\hat{\mathbf{R}}_k$, thereby increasing the robustness of this data-dependent beamformer for coherent sources (see [160, 102] for a detailed discussion). The purpose of using focusing matrices is to map the signal space at all frequency bins f_k from the range $[f_1, f_K]$ to a common reference frequency $f_0 \in [f_1, f_K]$, i.e.,

$$\mathbf{T}(\theta, f_k)\mathbf{a}(\theta, f_k) = \mathbf{a}(\theta, f_0).$$

If the Directions Of Arrival (DOAs) are known in the given scenario, the unitary focusing matrices $\mathbf{T}(\theta, f_k)$ [75] can be applied

$$\mathbf{T}(\theta, f_k) = Diag\left(\left[\frac{a_1(\theta, f_0)}{a_1(\theta, f_k)}, \frac{a_2(\theta, f_0)}{a_2(\theta, f_k)}, \ldots, \frac{a_M(\theta, f_0)}{a_M(\theta, f_k)}\right]\right), \tag{C.5}$$

where $Diag([\cdot])$ denotes the diagonal matrix with the elements of $[\cdot]$ on the main diagonal and $a_m(\theta, f_k)$ denotes the mth element of the steering vector $\mathbf{a}(\theta, f_k)$. Using (C.5), the focused and frequency smoothed measured autocorrelation matrix is

finally obtained as

$$\hat{\mathbf{R}} = \sum_{k=1}^{K} \mathbf{T}(\theta, f_k)\hat{\mathbf{R}}_k\mathbf{T}^H(\theta, f_k) \tag{C.6}$$

$$= \mathbf{a}(\theta, f_0)\hat{\mathbf{R}}_S\mathbf{a}^H(\theta, f_0) + \hat{\mathbf{R}}_Q, \tag{C.7}$$

where $\hat{\mathbf{R}}_Q = \sum_{k=1}^{K} \mathbf{T}(\theta, f_k)\hat{\mathbf{R}}_Q(f_k)\mathbf{T}^H(\theta, f_k)$ and $\hat{\mathbf{R}}_S = \sum_{k=1}^{K} \hat{\mathbf{R}}_S(f_k)$, where $\hat{\mathbf{R}}_S(f_k)$ and $\hat{\mathbf{R}}_Q(f_k)$ denote the source and noise power spectral densities, respectively. Substituting (C.6) into (C.3) and (C.1), the measured MVDR pseudospectrum can be estimated.

The MVDR design offers high directivity even at low frequencies and automatic null placement, which is beneficial for automatic suppression of the direct path and strong reflection signals. However, as a signal-dependent superdirective beamformer, the robustness issues mentioned above need to be addressed in order to deal with imperfect real microphone arrays and coherent reflection signals.

Appendix D

Least-squares beamshaping based on singular values truncation

In this Appendix we review the beamshaping technique presented in [111]. The acoustic beams (whose specifications are computed via beam tracing) are rendered by an array of loudspeakers using a least-squares (LS) approach with reference to a rendering equation that specifies the wavefield profile over a prescribed listening area. The spatial filters that produce the loudspeaker signals are obtained from the rendering equation through SVD.

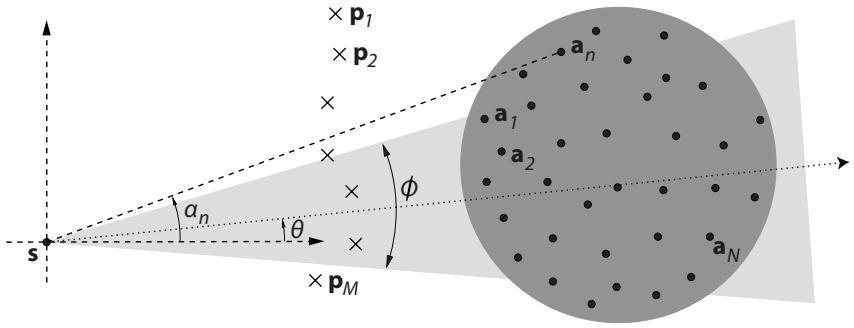


Figure D.1.: Geometry of the proposed rendering system.

According to Figure D.1, a set of ideal omnidirectional loudspeakers (i.e., point sources) are placed at arbitrary positions $\mathbf{p}_1, \ldots \mathbf{p}_M$ in an anechoic room. A set

of control points $\mathbf{a}_1, \dots \mathbf{a}_N$ is defined within the listening area that is depicted as the grey-shaded circle. The goal is to reproduce the acoustic beam generated by a virtual source located at $\mathbf{s}$ emitting towards the direction θ and with angular aperture ϕ. The acoustic source is constrained to be outside the listening area (note that some techniques for the rendering of sources surrounded by loudspeaker arrays have appeared in the literature [161, 162]). The Fourier transform $P_d(\omega, \mathbf{a}_n)$ of the desired wave field at the nth control point is therefore given by

$$P_d(\omega, \mathbf{a}_n) = g_\omega(\mathbf{a}_n|\mathbf{s})\Theta(\theta, \phi, \alpha_n)S(\omega) \ ,$$

where

$$g_\omega(\mathbf{a}_n|\mathbf{s}) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{s}-\mathbf{a}_n\|}}{4\pi\|\mathbf{s}-\mathbf{a}_n\|}$$

is the Green's function [131] from $\mathbf{s}$ to $\mathbf{a}_n$, ω being the frequency expressed in radians; and $\Theta(\theta, \phi, \alpha_n)$ is an angular function describing the directivity pattern of the beam. With reference to Figure D.1, α_n is the angle under which the virtual source is seen from $\mathbf{a}_n$; and $S(\omega)$ is the Fourier transform of the source signal. The goal of the beam shaping engine is to reproduce the effect of the virtual source at all the listening points by means of the loudspeakers. In other words, the aim is to find the vector of complex coefficients $\mathbf{h}_\omega$ applied to the loudspeakers, which satisfies the *rendering equation*

$$\mathbf{r}_\omega S(\omega) = \mathbf{G}_\omega \mathbf{h}_\omega S(\omega) \ , \tag{D.1}$$

where

$$\mathbf{G}_\omega = \begin{bmatrix} g_\omega(\mathbf{a}_1|\mathbf{p}_1) & \dots & g_\omega(\mathbf{a}_1|\mathbf{p}_M) \\ \vdots & \ddots & \vdots \\ g_\omega(\mathbf{a}_N|\mathbf{p}_1) & \dots & g_\omega(\mathbf{a}_N|\mathbf{p}_M) \end{bmatrix}$$

is the propagation matrix from each loudspeaker to each control point. The vector $\mathbf{r}_\omega$ encodes the desired response at all the control points, and it is therefore defined as

$$\mathbf{r}_\omega = [g_\omega(\mathbf{a}_1|\mathbf{s})\Theta(\theta, \phi, \alpha_1), \dots g_\omega(\mathbf{a}_N|\mathbf{s})\Theta(\theta, \phi, \alpha_N)]^T \ .$$

In other words, eq.(D.1) imposes that the spatial response of the loudspeaker array equals the desired soundfield. However, no exact solution exists in general, and the

best solution in the least-squares sense is given by

$$\hat{\mathbf{h}}_\omega = \mathbf{G}_\omega^+ \mathbf{r}_\omega = (\mathbf{G}_\omega^H \mathbf{G}_\omega)^{-1} \mathbf{G}_\omega^H \mathbf{r}_\omega \ .$$

A smooth beam-pattern can be obtained by choosing $N >> M$ and specifying $\Theta(\theta, \phi, \alpha_n)$ as a Gaussian function [8]. Unfortunately, in some cases the matrix $\mathbf{G}_\omega^H \mathbf{G}_\omega$ is ill-conditioned, and a reconditioning step is needed in order to provide feasible values to the coefficients $\hat{\mathbf{h}}_\omega$. As noticed in [111], a SVD-based reconditioning technique reveals to be suitable for given purposes. More sophisticated techniques exist, which obtain a better re-conditioning, such as those based on the Tikhonov regularization [163, 164], or based on the L1-norm minimization of the residuals of the rendering equation (Lasso technique) [136]. These techniques could guarantee an improved accuracy over the described one. However such methods are based on an approximation of the inverse of $\mathbf{G}_\omega$ and the inversion procedure depends on $\mathbf{r}_\omega$, thus requiring a demanding computation for every position of the virtual image sources and whenever the virtual source moves.

The spatial filter $\hat{\mathbf{h}}_\omega$ is sampled over a prescribed set of frequencies $\omega_1, ..., \omega_k, ..., \omega_K$ to obtain the matrix

$$\mathbf{H} = [\hat{\mathbf{h}}(\omega_1), \ldots, \hat{\mathbf{h}}(\omega_K)] \ ,$$

whose dimensions are $M \times K$. The mth row of $\mathbf{H}$ corresponds to the transfer function at frequencies $\omega_1, \ldots, \omega_K$ to be applied to the mth speaker in the array in order to render the presence of the desired acoustic beam. On the matrix $\mathbf{H}$ an interpolation is applied on the frequency axis in order to obtain the transfer function over a new and more densely populated set of frequencies $\omega_1, \ldots, \omega_{K'}$, $K' > K$. More specifically, a parabolic interpolation on the amplitude and a cubic interpolation on the phase are accomplished. The matrix resulting from this interpolation is $\mathbf{H}_{K'}$. The IFFT is applied on each row of $\mathbf{H}_{K'}$ to obtain the discrete time impulse responses $\hat{h}_m(t), m = 1, \ldots, M$, which are the filters to be applied to the loudspeakers in order to render the desired beam.

Bibliography

[1] P. Annibale, F. Antonacci, P. Bestagini, A. Brutti, A. Canclini, L. Cristoforetti, E. Habets, J. Filos, W. Kellermann, K. Kowalczyk, A. Lombard, E. Mabande, D. Marković, P. Naylor, M. Omologo, R. Rabenstein, A. Sarti, P. Svaizer, and M. Thomas, "The SCENIC project: Space-time audio processing for environment-aware acoustic sensing and rendering," in *Proc. of the 131st Audio Eng. Soc. Conv.*, Oct. 2011.

[2] A.A. Kaufman and A.L. Levshin, *Acoustic and Elastic Wave Fields in Geophysics, III*, Methods in Geochemistry and Geophysics. Elsevier Science, 2005.

[3] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *to appear on IEEE Transactions on Audio, Speech and Language Processing*, 2012.

[4] J. Filos, E. Habets, and P. Naylor, "A two-step approach to blindly infer room geometries," in *proc. of IEEE International Workshop on Acoustic Echo and Noise Cancellation (IWAENC'10)*, 2010.

[5] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 2620 –2623.

[6] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2012.

[7] M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1964 –1975, sept. 2012.

[8] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA '09*, New Paltz, New York, USA, October 2009, pp. 313 –316.

[9] D. Marković, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Visibility-based beam tracing for soundfield rendering," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, oct. 2010, pp. 40 – 45.

[10] F. Antonacci, A. Sarti, and S. Tubaro, "Two-dimensional beam-tracing from visibility diagrams for real-time acoustic rendering," *EURASIP Journal on Advances in Signal Processing*, vol. 1, pp. 1–18, 2010.

[11] A. Krokstad, S. Strom, 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, 1968.

[12] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1979.

[13] Michael Monks, Byong Mok Oh, and Julie Dorsey, "Acoustic simulation and visualization using a new unified beam tracing and image source approach," in *Proceedings of the 101th Audio Engineering Society Convention*, Los Angeles, California, USA, November 1996.

[14] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West, "A beam tracing approach to acoustic modeling for interactive virtual environments," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques, ACM, SIGGRAPH '98*, Orlando, Florida, USA, July 1998, pp. 21–32.

[15] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *Signal Processing, IEEE Transactions on*, vol. 54, no. 10, pp. 3790 –3804, oct. 2006.

[16] A. Kludszuweit, "Time iterative boundary element method (tibem) - a new numerical method of four-dimensional system analysis for the calculation of the spatial impulse response," *Acustica*, vol. 75, pp. 17–27, 1991.

[17] R.D. Ciskowski and C.A. Brebbia, *Boundary Element Methods in Acoustics*, International Series on Computational Engineering. Springer, 1991.

[18] S. Kopuz and N. Lalor, "Analysis of interior acoustic fields using the finite element method and the boundary element method," *Applied Acoustics*, vol. 45, no. 3, pp. 193 – 210, 1995.

[19] S Siltanen, T Lokki, and L Savioja, "Rays or waves? : Understanding the strengths and weaknesses of computational room acoustics modeling techniques," in *Proc. Int. Symposium on Room Acoustics*, Melbourne, Australia, Aug. 2010.

[20] J. Vian and D. V. Maercke, "Calculation of the room impulse response using a ray tracing method," *Proc. ICA Symp. Acoust. Theater Planning Performing Arts*, pp. 74–78, 1986.

[21] A.S. Glassner, *An Introduction to Ray Tracing*, Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. Academic Press, 1989.

[22] Eric Veach and Leonidas J. Guibas, "Metropolis light transport," in *Computer Graphics (SIGGRAPH '97 Proceedings*. 1997, pp. 65–76, Addison Wesley.

[23] Jeffrey Borish, "Extension of the image model to arbitrary polyhedra," *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[24] U.R. Kristiansen, A. Krokstad, and T. Follestad, "Extending the image method to higher-order reflections," *Applied Acoustics*, vol. 38, no. 2-4, pp. 195 – 206, 1993.

[25] John P. Walsh and Norman Dadoun, "What are we waiting for? the development of Godot, II," *Proceedings of the 103rd Meeting of the Acoustical Society of America*, vol. 71, no. S1, pp. S5, April 1982.

[26] Paul S. Heckbert and Pat Hanrahan, "Beam tracing polygonal objects," in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques, ACM, SIGGRAPH '84*, Minneapolis, Minnesota, USA, 1984, vol. 18, pp. 119–127.

[27] Norm Dadoun, David G. Kirkpatrick, and John P. Walsh, "The geometry of beam tracing," in *Proceedings of the first annual symposium on Computational geometry, ACM, SCG '85*, Sedona, Arizona, USA, June 1985, pp. 55–61.

[28] U. Stephenson and U. Kristiansen, "Pyramidal beam tracing and time dependent radiosity," in *Proceedings of the 15th International Congress on Acoustics*, Trondheim, Norway, June 1995, pp. 657–660.

[29] Thomas Funkhouser, Patrick Min, and Ingrid Carlbom, "Real-time acoustic modeling for distributed virtual environments," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM, SIGGRAPH '99*, Los Angeles, California, USA, 1999, pp. 365–374, ACM Press/Addison-Wesley Publishing Co.

[30] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," *Journal of the Acoustical Society of America*, vol. 115, pp. 739–756, 2004.

[31] Samuli Laine, Samuel Siltanen, Tapio Lokki, and Lauri Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172 – 181, 2009.

[32] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 4, pp. 812 –824, may 2008.

[33] Edward H. Adelson and James R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. 1991, pp. 3–20, MIT Press.

[34] Jorge Stolfi, *Oriented Projective Geometry: A Framework for Geometric Computations*, Academic Press, 1991.

[35] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoutic emissions," in *proc. of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing , ICASSP-2010, Dallas, TX*, March 14-19 2010, pp. pp. 2822–2825.

[36] D. Aprea, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic reconstruction of the geometry of an environment through acquisition of a controlled emission," in *Proc. of EUSIPCO*, 2009.

[37] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 153 –156.

[38] I. Dokmanic, Y.M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-D polygonal case," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 321 –324.

[39] F. Ribeiro, D. Florencio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2012.

[40] E. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of planar acoustic reflectors through emission of controlled stimuli," in *19th European Signal Processing Conference (EUSIPCO 2011)*, 2011.

[41] J. Filos, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of planar reflectors from the combination of linear estimates," in *proc. of European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 1019–1023.

[42] Jiri Bittner, *Hierarchical Techniques for Visibility Computations*, Ph.D. thesis, Czech Technical University in Prague, Prague, Czech Republic, 2002.

[43] Jiri Bittner, Jan Prikryl, and Pavel Slavik, "Exact regional visibility using line space partitioning," *Computers & Graphics*, vol. 27, no. 4, pp. 569 – 580, 2003.

[44] D.M.L.Y. Sommerville, *Analytical Geometry of Three Dimensions*, The University Press, 1934.

[45] B. Chazelle, H. Edelsbrunner, L. J. Guibas, M. Sharir, and J. Stolfi, "Lines in space: Combinatorics and algorithms," *Algorithmica*, vol. 15, pp. 428–447, 1996, 10.1007/BF01955043.

[46] Seth J Teller, *Visibility Computations in Densely Occluded Polyhedral Environments*, Ph.D. thesis, University of California at Berkeley, Berkeley, CA, USA, 1992.

[47] Seth J Teller and Michael E Hohmeyer, "Computing the lines piercing four lines," Tech. Rep., Berkeley, CA, USA, 1992.

[48] Seth Teller and Michael Hohmeyer, "Determining the lines through four lines," *J. Graph. Tools*, vol. 4, no. 3, pp. 11–22, November 1999.

[49] Seth Teller and Michael Hohmeyer, "Stabbing oriented convex polygons in randomized $o(n^2)$ time," in *In CONM Vol. 178: Proc. Jerusalem Combinatorics '93*. 1994, pp. 311–318, American Mathematical Society.

[50] Seth J. Teller, "Computing the antipenumbra of an area light source," *SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 139–148, July 1992.

[51] Vladlen Koltun, Yiorgos Chrysanthou, and Daniel Cohen-Or, "Hardware-accelerated from-region visibility using a dual ray space," in *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*, London, UK, UK, 2001, pp. 205–216, Springer-Verlag.

[52] Jiri Bittner and Peter Wonka, "Visibility in computer graphics," *Journal of Environmental Planning*, vol. 30, pp. 729–756, 2003.

[53] A. Canclini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround systesm exploiting early reflections in a reverberant room," in *proc. of European Signal Processing Conference (EUSIPCO)*, 2012.

[54] Seth J. Teller and Carlo H. Séquin, "Visibility preprocessing for interactive walkthroughs," *SIGGRAPH Comput. Graph.*, vol. 25, no. 4, pp. 61–70, July 1991.

[55] Seth Teller, Celeste Fowler, Thomas Funkhouser, and Pat Hanrahan, "Partitioning and ordering large radiosity computations," 1994.

[56] Tobias Schröder, Dirk; Lentz, "Real-time processing of image sources using binary space partitioning," *J. Audio Eng. Soc*, vol. 54, no. 7/8, pp. 604–619, 2006.

[57] Christian Lauterbach, Anish Chandak, and Dinesh Manocha, "Interactive sound rendering in complex and dynamic scenes using frustum tracing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1672–1679, November 2007.

[58] Anish Chandak, Christian Lauterbach, Micah T. Taylor, Zhimin Ren, and Dinesh Manocha, "Ad-frustum: Adaptive frustum tracing for interactive sound propagation," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1707–1722, 2008.

[59] F. Durand, *3D visibility: Analytical study and applications*, Ph.D. thesis, Univ. Joseph Fourier, Grenoble, France, 1999.

[60] Daniel Cohen-Or, Yiorgos L. Chrysanthou, Cláudio T. Silva, and Frédo Durand, "A survey of visibility for walkthrough applications," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 9, no. 3, pp. 412 – 431, july-sept. 2003.

[61] S. Nirenstein, E. Blake, and J. Gain, "Exact from-region visibility culling," in *Proceedings of the 13th Eurographics workshop on Rendering*, Aire-la-Ville, Switzerland, Switzerland, 2002, EGRW '02, pp. 191–202, Eurographics Association.

[62] Shaun Nirenstein, *Fast and Accurate Visibility Preprocessing*, Ph.D. thesis, University of Cape Town, 2003.

[63] Denis Haumont, Otso Makinen, and Shaun Nirenstein, "A low dimensional framework for exact polygon-to-polygon occlusion queries," in *Rendering Techniques*, 2005, pp. 211–222.

[64] F. Mora, L. Aveneau, and M. Mériaux, "Coherent and exact polygon-to-polygon visibility.," in *WSCG'2005, Plzen, République Tchèque.*, Janvier 2005.

[65] F. Mora and L. Aveneau, "Fast and exact direct illumination," in *Proceedings of the Computer Graphics International 2005*, Washington, DC, USA, 2005, CGI '05, pp. 191–197, IEEE Computer Society.

[66] Sylvain Charneau, Lilian Aveneau, and Laurent Fuchs, "Exact, robust and efficient full visibility computation in P lucker space," *Vis. Comput.*, vol. 23, no. 9, pp. 773–782, August 2007.

[67] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004.

[68] George Drettakis and Eugene Fiume, "A fast shadow algorithm for area light sources using backprojection," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1994, SIGGRAPH '94, pp. 223–230, ACM.

[69] Frédo Durand, George Drettakis, and Claude Puech, "The visibility skeleton: a powerful and efficient multi-purpose global visibility tool," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1997, SIGGRAPH '97, pp. 89–100, ACM Press/Addison-Wesley Publishing Co.

[70] J. Berent and P.L. Dragotti, "Plenoptic manifolds," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 34 –44, nov. 2007.

[71] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen, "The lumigraph," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, SIGGRAPH '96, pp. 43–54, ACM.

[72] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, SIGGRAPH '96, pp. 31–42, ACM.

[73] Ren Ng, *Digital light field photography*, Ph.D. thesis, Stanford, CA, USA, 2006, AAI3219345.

[74] Edward H. Adelson and John Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, February 1992.

[75] M. R. Azimi-Sadjadi, A. Pezeshki, L. L. Scharf, and M. Hohil, "Wideband DOA estimation algorithms for multiple target detection and tracking using unattended acoustic sensors," *Proc. of the SPIE04 Defense and Security Symposium*, vol. 5417, pp. 111, April 2004.

[76] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, chapter 6: Spatial Methods, pp. 232–238, Prentic, 1997.

[77] H.L. van Trees, *Optimum array processing: Detection, estimation, and modulation theory*, Wiley, New York, 2002.

[78] E. Mabande, Haohai Sun, K. Kowalczyk, and W. Kellermann, "On 2D localization of reflectors using robust beamforming techniques," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 153 –156.

[79] H.L. Van Trees and J. Wiley, *Optimum array processing*, chapter 2, pp. 17–89, Wiley–Interscience, 2002.

[80] E. Fisher and B. Rafaely, "Near-field spherical microphone array processing with radial filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 256–265, Feb. 2011.

[81] Haohai Sun, E. Mabande, K. Kowalczylc, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 113 –116.

[82] D. Ba, F. Ribeiro, Cha Zhang, and D. Florencio, "L1 regularized room modeling with compact microphone arrays," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 157 –160.

[83] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.

[84] J. Filos, A. Canclini, M. Thomas, F. Antonacci, A. Sarti, and P. Naylor, "Robust inference of room geometry from acoustic measurements using the hough transform," in *19th European Signal Processing Conference (EUSIPCO 2011)*, 2011.

[85] 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 *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, oct. 2011, pp. 17 –20.

[86] A. Canclini, F. Antonacci, J. Filos, A. Sarti, and P. A. Naylor, "Exact localization of planar acoustic reflectors in three-dimensional geometries," in *International Workshop on Acoustic Signal Enhancement (IWAENC 2012)*, RWTH Aachen University, Sept. 2012.

[87] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 10, pp. 1365 – 1376, oct 1987.

[88] T.D. Abhayapala and H. Bhatta, "Coherent broadband source localization by modal space processing," in *Telecommunications, 2003. ICT 2003. 10th International Conference on*, feb.-1 march 2003, vol. 2, pp. 1617 – 1623 vol.2.

[89] E. Mabande, A. Schad, and W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, april 2009, pp. 77 –80.

[90] Boaz Rafaely, Yotam Peled, Morag Agmon, Dima Khaykin, and Etan Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication*, Israel Cohen, Jacob Benesty, and Sharon Gannot, Eds., vol. 3 of *Springer Topics in Signal Processing*, pp. 281–305. Springer Berlin Heidelberg, 2010.

[91] P. Handel, P. Stoica, and T. Soderstrom, "Capon method for DOA estimation: accuracy and robustness aspects," in *Nonlinear Digital Signal Processing, 1993. IEEE Winter Workshop on*, 1993.

[92] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Number v. 1 in Prentice Hall Signal Processing Series. Prentice-Hall PTR, 1993.

[93] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[94] P. Teng, A. Lombard, and W. Kellermann, "Disambiguation in multidimensional tracking of multiple acoustic sources using a gaussian likelihood criterion," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 145–148.

[95] P. V. C. Hough, "Method and means for recognizing complex patterns," December 1962.

[96] Richard O. Duda and Peter E. Hart, "Use of the H ough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, January 1972.

[97] Jean-Michel Jolion, Peter Meer, and Samira Bataouche, "Robust clustering with applications in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 791–802, August 1991.

[98] Yasutaka Furukawa and Yoshihisa Shinagawa, "Accurate and robust line segment extraction by analyzing distribution around peaks in hough space," *Computer Vision and Image Understanding*, vol. 92, no. 1, pp. 1 – 25, 2003.

[99] Y. Peled and B. Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, pp. 113–116, Mar. 2010.

[100] M. Kuster, "Multichannel room impulse response rendering on the basis of underdetermined data," *J. Audio Eng. Soc.*, vol. 157, no. 6, pp. 403–412, 2009.

[101] S. Tervo, T. Korhonen, and T. Lokki, "Estimation of reflections from impulse responses," *Proc. Int. Symp. Room Acoustics*, pp. 1–7, Aug. 2010.

[102] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. 2828–2840, 2012.

[103] M. Kuster, "Reliability of estimating the room volume from a single room impulse response," *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 982–993, 2008.

[104] H. Chung, H. Shim, N. Hahn, S.B. Chon, and K.M Sung, "Sound reproduction method by front loudspeaker array for home theater applications," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 528 –534, 2012.

[105] H. Kuttruff, *Room acoustics*, 4th Ed. Spon Press, London, 2000.

[106] E. Mommertz, "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," *Applied Acoustics*, vol. 46, pp. 251–265, 1995.

[107] J. Ducourneau, V. Planeau, and A. Nejade, "Design of a multipolar weighting for acoustic antennae," *Applied Acoustics*, vol. 70, pp. 484–492, 2009.

[108] A. Brutti, M. Omologo, and P. Svaizer, "Inference of acoustic source directivity using environment awareness," in *proc. of EURASIP Signal Processing Conference, EUSIPCO*, 2012.

[109] P. C. Meuse and H. F. Silverman, "Characterization of talker radiation pattern using a microphone array," in *proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1994, pp. 257 – 260.

[110] M. Feder, "Parameter estimation of superimposed signals using the em algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 4, pp. 477 – 489, Apr 1988.

[111] Fabio Antonacci, Alberto Calatroni, Antonio Canclini, Andrea Galbiati, Augusto Sarti, and Stefano Tubaro, "Rendering of an acoustic beam through an array of loudspeakers," in *Proceedings of the 12th International Conference on Digital Audio Effects, DAFx-09*, Como, Italy, September 2009.

[112] Antonio Canclini, *A geometric approach to analysis and synthesis of acoustic wave fields for multimedia applications*, Ph.D. thesis, Politecnico di Milano, 2012.

[113] J. Ahrens, *Analytic Methods of Sound Field Synthesis*, Tlabs Series in Telecommunication Services. Springerverlag Berlin Heidelberg, 2012.

[114] A. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J.Acoust.Soc.Am.*, vol. 93, pp. 2764–2778, 1993.

[115] Jens Ahrens, Rudolph Rabenstein, and Sascha Spors, "The theory of wave field synthesis revisited," in *Audio Engineering Society Convention 124*, May 2008.

[116] Jerome Daniel, Sebastien Moreau, and Rozenn Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, 3 2003.

[117] Rozenn Nicol, "Sound spatialization by higher order ambisonics: Encoding and decoding a sound scene in practice from a theoretical point of view.," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, 2010.

[118] Zhaoxi Wang and S. F. Wu, "Helmholtz equation – least-squares method for reconstructing the acoustic pressure field," *The Journal of the Acoustical Society of America*, vol. 102, no. April, pp. 2020, 1997.

[119] Philippe-Aubert Gauthier, Alain Berry, and Wieslaw Woszczyk, "Sound-field reproduction in-room using optimal control techniques: simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 117, no. 2, pp. 662–78, 2005.

[120] D.B. Ward and T.D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, sep 2001.

[121] L.D. Fielder, "Practical limits for room equalization," in *Audio Engineering Society Convention 111*, 11 2001.

[122] J.J. Lopez, A. Gonzalez, and L. Fuster, "Room compensation in wave field synthesis by means of multichannel inversion," in *2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, oct. 2005, pp. 146 – 149.

[123] S. Spors, A. Kuntz, and R. Rabenstein, "An Approach to Listening Room Compensation with Wave Field Synthesis," in *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 2003, pp. 49–52.

[124] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J Acoust Soc Am*, vol. 122, pp. 354–369, 2007.

[125] M. Schneider and W. Kellermann, "Adaptive Listening Room Equalization Using a Scalable Filtering Structure in the wave domain," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012.

[126] R.S. Pellegrini, "Perception-based room rendering for auditory scenes," in *Audio Engineering Society Convention 109*, Sep. 2000.

[127] J. Pope, D. Creasey, and A. Chalmers, "Realtime room acoustics using ambisonics," in *Audio Engineering Society Conference, 16th International Conference: Spatial Sound Reproduction*, Mar. 1999.

[128] A. Wabnitz, N. Epain, A. van Schaik, and C. Jin, "Time domain reconstruction of spatial sound fields using compressed sensing," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 465 –468.

[129] M. Kahrs and K. Brandenburg, *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer International Series in Engineering and Computer Science. Kluwer, 1998.

[130] S. Spors, *Active Listening Room Compensation for Spatial Sound Reproduction Systems*, Technische Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg., 2006.

[131] Earl G. Williams, *Fourier acoustics : sound radiation and nearfield acoustical holography*, Academic Press, June 1999.

[132] B. Pueo, J. Escolano, and M. Roma, "Precise control of beam direction and beamwidth of linear loudspeaker arrays," in *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004*, july 2004, pp. 538 – 541.

[133] J. Atkins, "Robust beamforming and steering of arbitrary beam patterns using spherical arrays," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, oct. 2011, pp. 237–240.

[134] Yoomi Hur, Seong Woo Kim, Young-cheol Park, and Dae Hee Youn, "Highly focused sound beamforming algorithm using loudspeaker array system," in *Audio Engineering Society Convention 125*, 10 2008.

[135] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 320 –328, apr 1994.

[136] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.

[137] B.C.J. Moore, *Hearing*, Handbook of Perception and Cognition. Academic Press, 1995.

[138] B.D. Radlović, R.C. Williamson, and R.A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 311–319, 2000.

[139] L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, and S. Tubaro, "A psychoacoustic-based analysis of the impact of pre-echoes and post-echoes in soundfield rendering applications," in *International Workshop on Acoustic Signal Enhancement (IWAENC 2012)*, September 2012.

[140] L.L. Beranek and Acoustical Society of America, *Acoustics*, Electrical and electronic engineering. American Institute of Physics, 1986.

[141] D.T. Blackstock, *Fundamentals of Physical Acoustics*, Wiley-Interscience. Wiley, 2000.

[142] P.M.C. Morse and H. Feshbach, *Methods of theoretical physics*, Number v. 1 in International series in pure and applied physics. McGraw-Hill, 1953.

[143] T.D. Rossing, *Springer Handbook of Acoustics*, Springer Handbook of. Springer, 2007.

[144] V. Troyan and Y. Kiselev, *Statistical Methods of Geophysical Data Processing*, World Scientific, 2010.

[145] Monin, *Theoretical Geophysical Fluid Dynamics*, Environmental Fluid Mechanics. Springer, 1990.

[146] C. Goral, K. Torrance, D. Greenberg, and B.Battaile, "Modeling the interaction of light between diffuse surfaces," in *Proceedings of ACM Computer Graphics, SIGGRAPH 84, Minneapolis, Minnesota*, 1984, vol. 18, pp. 213–222.

[147] G.R.Moore, *An approach to the Analysis of Sound in Auditoria*, Ph.D. thesis, Cambridge, 1984.

[148] Michael F. Cohen, John Wallace, and Pat Hanrahan, *Radiosity and realistic image synthesis*, Academic Press Professional, Inc., San Diego, CA, USA, 1993.

[149] T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics," *Applied Acoustics*, vol. 38, no. 2-4, pp. 161 – 178, 1993.

[150] N. Tsingos and J.D. Gascuel, "A general model for simulation of room acoustics based on hierarchical radiosity," in *Visual Proceedings of ACM Computer Graphics, SIGGRAPH 97, Los Angeles*, 1997.

[151] Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom, "Modeling acoustics in virtual environments using the uniform theory of diffraction," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 2001, SIGGRAPH '01, pp. 545–552, ACM.

[152] R.G. Kouyoumjian and P.H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proceedings of the IEEE*, vol. 62, no. 11, pp. 1448 – 1461, nov. 1974.

[153] Joseph B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Am.*, vol. 52, no. 2, pp. 116–130, 1962.

[154] M. Kaveh and A. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *Acoustics,*

Speech and Signal Processing, IEEE Transactions on, vol. 34, no. 2, pp. 331 – 341, apr 1986.

[155] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 4, pp. 532 –544, apr 1988.

[156] P. Stoica and Nehorai Arye, "MUSIC, maximum likelihood, and Cramer-Rao bound," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 5, pp. 720 –741, may 1989.

[157] B.D. Rao and K.V.S. Hari, "Performance analysis of root-MUSIC," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 12, pp. 1939 –1949, dec 1989.

[158] B.D. Van Veen and K.M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, April 1988.

[159] J. Bitzer and K.U. Simmer, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 2: Superdirective microphone arrays, pp. 19–37, Springer, Berlin, 2001.

[160] T. D. Abhayapala, *Advances in Direction-of-Arrival Estimation*, chapter Broadband Source Localization by Modal Space Processing, Artech House, 2006.

[161] Jung-Woo Choi and Yang-Hann Kim, "Integral approach for reproduction of virtual sound source surrounded by loudspeaker array," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1976 –1989, sept. 2012.

[162] S. Spors and J. Ahrens, "Reproduction of focused sources by the spectral division method," in *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, march 2010, pp. 1 –5.

[163] Y. Kim and P.A. Nelson, "Optimal regularisation for acoustic source reconstruction by inverse methods," *Journal of Sound and Vibration*, vol. 275, no. 35, pp. 463 – 487, 2004.

[164] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2012.