Politecnico di Milano
Facoltà di Ingegneria dei sistemi

Corso di Laurea in Ingegneria matematica
Indirizzo Scienze Computazionali per l'Ingegneria

Acoustic Imaging in the RaySpace: Application to Environment Inference

Tesi di secondo livello di:
Giorgio Sandrini

Relatore:
Prof. Augusto Sarti - Politecnico di Milano
Correlatore:
Dr. Dejan Markovic - Politecnico di Milano

Matricola 750097
Sessione di Laurea Aprile 2012
Anno Accademico 2011–2012

"Strada: striscia di terra che si percorre a piedi. Diversa dalla strada è la strada asfaltata, che si distingue non solo perché la si percorre con la macchina, ma in quanto è una semplice linea che unisce un punto a un altro. La strada asfaltata non ha senso in se stessa; hanno senso solo i due punti che essa unisce. La strada è una lode allo spazio. Ogni tratto di strada ha senso in se stesso e ci invita alla sosta. La strada asfaltata è una trionfale svalutazione dello spazio, che per suo merito oggi non è che un semplice ostacolo al movimento dell'uomo e una perdita di tempo. Prima ancora di scomparire dal paesaggio, le strade sono scomparse dall'animo umano: l'uomo ha smesso di desiderare di camminare con le proprie gambe e di gioire per questo. Anche la propria vita ormai non la vede più come una strada, bensì come una strada asfaltata: come una linea che conduce da un punto a un altro, dal grado di capitano al grado di generale, dal ruolo di moglie al ruolo di vedova. Il tempo della vita è diventato per lui un semplice ostacolo che è necessario superare a velocità sempre maggiore"

M. Kundera, L'immortalità, p. 242-243.

Abstract

This work is part of the environment-aware processing, a field of recent interest that promises to significantly push the boundaries of audio signal processing. The environment-aware processing uses information arising from the environmental response to enable the acoustic systems to become aware of their own characteristics and geometry and those of the environment that they operate in. This information allows advanced and innovative space-time processing solutions.

In particular the thesis addresses the problem of inference from acoustic measurements on the geometric characteristics of the environment. Recently a number of techniques for localization of reflective surfaces appeared in literature. These techniques exploit different information extracted from acoustic measurements to infer the position of the reflectors in the environment. Usually the extracted information, combined with some a priori knowledge, defines a non-linear constraint on reflector position. Using multiple constraints (e.g changing the hardware position) a cost function is formulated whose minimization yields the estimated line or plane (for 2D or 3D geometries) on which the reflector lies.

In this work we take a slightly different approach for the localization of reflective surfaces. Instead of extracting information related to a specific geometric constraint, we are interested in "looking" at the acoustic scene, i.e. obtaining an overview of what is happening in different positions in space, and successively estimating the environment geometry from a number of such acoustic "snapshots". Therefore, we want to imitate, to a certain extent, the procedures used in computer vision to reconstruct the environment geometry taking visual snapshots from different points of view.

The acoustical snapshots are defined using a non-linear transformation applied to acoustic measurements that maps the data in a space in which the

geometric primitives are represented by linear constraints.

Unlike most of other methods, the acoustic observation of the environment allows us to find not only the line on which the reflector lies but also its extension. This property can turn useful in irregular, complex environments where occlusions and limited visibility of acoustic reflectors are present.

Furthermore, the representation of acoustic measurements defined in this work can potentially be used also to infer on radiometric properties of the environment (e.g. radiation pattern and reflection coefficients) and therefore it has a number of potential other applications.

Sommario

Le applicazioni avanzate di elaborazione dei segnali acustici, in particolare quelle basate su array di sensori, sono fortemente sensibili alla risposta acustica dell'ambiente. Il riverbero sonoro generato in un ambiente confinato è solitamente visto da tali tecniche come un fattore di disturbo che è necessario compensare, mentre la natura ci insegna come la risposta ambientale a stimoli acustici è fonte di informazioni fondamentali sulle caratteristiche dell'ambiente stesso. Trasformare la risposta acustica da un fattore di errore in una risorsa richiede una profonda comprensione dei fenomeni di propagazione ed una accurata modellizzazione acustica dell'ambiente. Questo può essere fatto eccitando l'ambiente attraverso emissioni sonore generate da fonti che mostrino una struttura temporale e spaziale.

L'environment-aware processing permette ai sistemi acustici di diventare consapevoli delle proprie caratteristiche e di quelle dell'ambiente in cui operano. Esso consente avanzate soluzioni spazio-temporali di elaborazione che sfruttano le informazioni aggiuntive fornite dalla risposta acustica dell' ambiente di interesse. Tale tipo di approccio promette di far avanzare in modo significativo i confini della elaborazione dei segnali audio. In particolare, tali tecniche possono essere utilizzate per aumentare le prestazioni di algoritmi per la localizzazione e caratterizzazione delle sorgenti in ambienti riverberanti, per migliorare gli algoritmi di compensazione del riverbero, per eseguire il rendering dell'ambiente virtuale etc.

Un punto essenziale comune alle tecniche di environment-aware processing è l'utilizzo simultaneo di sensori e fonti sonore allo scopo di inferire sulle caratteristiche dell'ambiente. Recentemente, numerose tecniche per la localizzazione delle superfici riflettenti sono apparse in letteratura. Tali tecniche sfruttano informazioni estratte da misurazioni acustiche per stimare la posizione delle pareti nell'ambiente. Normalmente le informazione estratte, in

combinazione con qualche assunzione a priori, determinano la definizione di un vincolo non-lineare sulla posizione del riflettore. Combinando più vincoli (ad esempio per posizioni di sorgenti o microfoni differenti) una funzione di costo è definita e la stima della linea corrispondente al riflettore è ottenuta attraverso una procedura di minimizzazione. I metodi proposti generalmente differiscono per il tipo di hardware richiesto e per le assunzioni a priori definite e possono essere più adatti a tipi differenti di scenario.

In questo lavoro affrontiamo il problema della inferenza sulla geometria dell'ambiente partendo da un approccio differente. Invece di estrarre informazioni relative ad un vincolo geometrico specifico, siamo interessati a "guardare" la risposta acustica nel suo complesso, vale a dire ottenere una panoramica su ciò che sta accadendo nello spazio al variare delle posizioni delle sorgenti, e successivamente stimare la geometria dell'ambiente da un numero di tali "immagini" acustiche. L'idea è quella di imitare le procedure utilizzate in computer vision per ricostruire la geometria dell'ambiente scattando istantanee visive da punti di vista differenti. Come nella computer vision si ha bisogno della luce per illuminare la scena, noi avremo bisogno di sorgenti acustiche per stimolare una risposta nell'ambiente. Tale risposta sarà parzialmente acquisita da un array di microfoni che rappresenta la nostra camera acustica.

Le misurazioni acustiche ottenute dall'array di microfoni saranno descritte con coordinate opportune. Analogamente alle immagini in ottica, in cui ogni pixel rappresenta un raggio passante per il corrispondente centro ottico, per il caso in esame mapperemo i dati a disposizione nello spazio dei raggi acustici attraverso una funzione non lineare. Tale rappresentazione delle misurazioni acustiche sarà denominata "immagine acustica". Diversamente dal caso relativo all'ottica, tuttavia, non è possibile costruire un dispositivo che catturi la magnitudo dei raggi provenienti da una data direzione direttamente. Per ottenere tali informazioni sarà necessario utilizzare tecniche di analisi spazio-temporali del segnale acquisito che introdurranno una distorsione rispetto all'immagine acustica ideale. Da diverse immagini acustiche così ottenute si prenderà in considerazione il problema dell'inferenza e della ricostruizione della geometria "illuminata".

Sebbene il metodo proposto in questo lavoro risulti in genere più esigente a livello hardware rispetto agli algoritmi pre-esistenti in letteratura (vedremo

che per "vedere" porzioni più grandi di spazio si richiede un array di microfoni lungo o in movimento ed una sorgente in movimento) l'osservazione acustica dell'ambiente consentirà, a differenza degli altri metodi, di stimare non solo la direzione dei riflettori ma anche la loro estensione. In particolare le estensioni saranno stimabili per riflettori i cui punti terminali sono visibili dall'array di microfoni. Tale proprietà risulterà molto utile in ambienti irregolari e complessi in cui occlusioni e visibilità limitata delle pareti influenzano notevolmente il campo acustico complessivo e devono essere prese in considerazione dai sistemi acustici in una ottica di environment-aware processing.

Inoltre, l'approccio proposto ha anche una serie di altre caratteristiche interessanti. Una trasformazione non lineare mappa le informazioni in uno spazio in cui le primitive geometriche sono rappresentate da vincoli lineari. Questo comporta che nessuna procedura iterativa di minimizzazione o soluzione complessa in forma chiusa è richiesta e le stime possono essere ottenute usando il metodo dei minimi quadrati. Inoltre, vedremo che non sarà necessario controllare le sorgenti dirette che illuminano la scena acustica. L'algoritmo può localizzare superfici riflettenti senza le informazioni sulla posizione di tali sorgenti. Inoltre, se visibili, le posizioni delle sorgenti dirette non solo non sono richieste come dato al problema ma possono essere inoltre stimate automaticamente anch'esse dalle immagini acquisite.

Le prestazioni degli algoritmi di stima proposti saranno analizzate nel senso delle proprietà statistiche del secondo ordine. Tale analisi sarà in primo luogo utile per la progettazione di sistemi con le caratteristiche di varianza-/covarianza desiderate per gli output del problema. Inoltre, essa sarà anche utilizzata per migliorare le prestazioni stesse degli algoritmi utilizzando una procedura iterativa che permetterà di approssimare una stima a massima verosimiglianza (ML) delle incognite.

Infine, le immagini acustiche introdotte potranno essere eventualmente utili in altre applicazioni. Sebbene in questo lavoro tali immagini vengano utilizzate per dedurre solo le informazioni geometriche, esse catturano infatti anche le caratteristiche radiometriche dell'ambiente (ad esempio i radiation pattern delle sorgenti e i coefficienti di riflessione delle pareti) quindi esse hanno molte potenziali ulteriori applicazioni, principalmente per quanto riguarda l'estrazione di informazioni dall'ambiente ma anche, per esempio la predizione della pressione acustica in un punto generico dello spazio.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The first chapter aims to introduce the problem addressed in this thesis and to present the motivation for this work. The concepts presented here will be formalized and explained in more detail in the following chapters.

Advanced sound processing applications, particularly those based on array processing, are critically sensitive to the environmental acoustic response because their design does not account for the complex propagation phenomena that cause it. Reverberations are usually seen as a liability to take countermeasures against, while nature teaches us that the information provided by the acoustic interaction with the environment can become a valuable asset that enables complex navigational tasks and more. Turning the acoustic response from a liability into an asset requires a thorough understanding of propagation phenomena, and an accurate acoustic modelling of the environment. This can be done by listening to how the environment renders sound emissions generated by sources, as long as such emissions exhibit a temporal as well as a spatial structure. These issues are addressed in the European project SCENIC (Self-Configuring ENvironment-aware Intelligent aCoustic sensing) [1].

The environment-aware processing enables acoustic systems to become aware of their own characteristics and geometry and those of the environment that they operate in. It allows advanced space-time processing solutions that take advantage of the additional information provided by the environmental acoustic response. That kind of approach to acoustical problems promises to significantly push the boundaries of audio signal processing. In fact, the environment-awareness can be used to boost the performance of algorithms

for localization and characterization of sources in reverberant environments [2] that are now able to predict the effect of reverberations on the soundfield and therefore to exploit reverberations to improve their accuracy. It results also important in order to improve dereverberation algorithms and to perform wavefield rendering [3] that, thanks to that kind of approach, can virtually knock out the walls of the environment in which it operates (i.e. the system removes the effect of reverberations) and renders the acoustics of a virtual environment.

In Figure 1.1, we show a flow diagram of the information that are used for the environment-aware processing. One key point is that, in order to achieve this status of awareness, sensors and sources have to be used together in a synergistic fashion. Inference is the first step in environment-aware signal processing. It is the study of the acoustic measurements for the estimation of the characteristics of an acoustic environment and the sources within it. The thesis deals with this part of the problem and, in particular, the localization of the reflective surfaces.

Motivated by the recent progress in environment-aware processing the problem of inference is becoming an important issue in audio signal processing. This problem generally considers two types of information: low-level features include the location of acoustic reflectors relative to a microphone array and the directivity of an acoustic source and high-level features include temperature and reverberation time. Recently a number of techniques for localization of reflective surfaces appeared in literature. These techniques exploit different information extracted from acoustic measurements to infer the position of the reflectors in the environment. Usually the extracted information, in combination with some a priori knowledge, defines a non-linear constraint on reflector position. Combining multiple constraints (e.g. for different source or microphone positions) a cost function is defined whose minimization yields the estimated line or plane (for 2D or 3D geometries) on which the reflector lies. Proposed methods generally achieve a considerable accuracy but differ in assumptions made and hardware used making some of them more suitable for certain scenarios than others.

In this work we take a slightly different approach for the problem of the localization of reflective surfaces. Instead of extracting information related to a specific geometric constraint, we are interested in "looking" at the acoustic

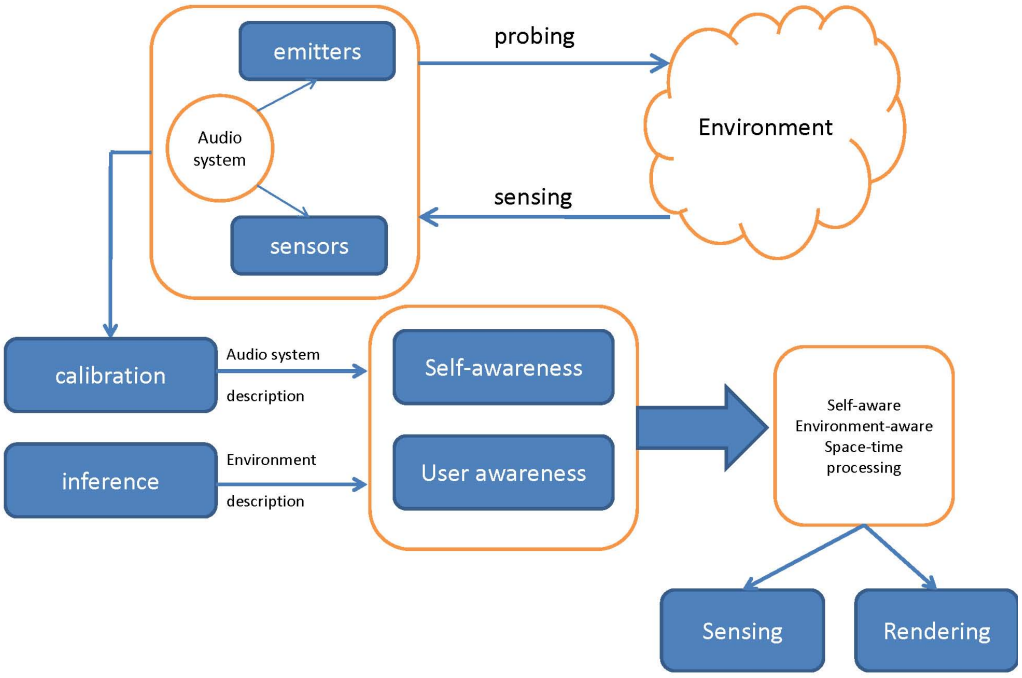


Figure 1.1: A flow diagram of the environment-aware processing: through the sound system, environment is stimulated and its response is acquired; From the information deriving from environment response you infer on the audio system and/or environment description; The self awareness and environment-awareness are used to the desired purpose.

scene, i.e. obtaining an overview of what is happening in different positions in space, and successively estimating the environment geometry from a number of such acoustic "snapshots". Therefore we are imitating, to a certain extent, the procedures used in computer vision to reconstruct the environment geometry taking visual snapshots from different points of view. As in computer vision we need the light to illuminate the scene, we will need acoustic sources to stimulate a response in the environment. That response will be partially acquired by a microphone array that represent our acoustic camera.

However, to do so we have to represent the acoustic measurements obtained by the microphone array with suitable coordinates. Similarly to images in optics, where each pixel represents a ray passing through the pixel point and the optical center, we map the acoustic measurements in the space of acoustic rays. Such representation of the acoustic measurements will be

referred here as the "acoustic image". Unlike optics, however, it is not possible to build a device that captures the magnitude of rays coming from a given direction. To obtain information about direction we use space-time processing techniques and in doing so we introduce distortion with respect to the ideal acoustic image. From several acoustical images obtained in the described manner, we want to take into account the inference problem and reconstruct the "illuminated" geometry.

Although generally more demanding with respect to dedicated algorithms for geometry inference (we will see that in order to "see" bigger portions of space it requires a long or moving array and a moving source), unlike most of other methods, the acoustic observation of the environment allows us to find not only the line on which the reflector lies but also its extension. In particular, the extension can be estimated if the reflector endpoints are acoustically visible from the microphone array. This property can turn useful in irregular, complex environments where occlusions and limited visibility of acoustic reflections greatly characterize the overall soundfield and have to be taken into account by the environment-aware systems operating inside such environments. Furthermore, the proposed approach has also a number of other interesting features. A non-linear transformation applied to acoustic measurements maps the data in the space in which the geometric primitives are represented by linear constraints. No iterative minimization procedures or complex solutions in a closed form are required and the solutions can be found using the least squares method. Moreover, we will see that the sources that are used have not necessarly to be controlled, the algorithm can localize reflective surfaces without the information of the position for the direct sources. Moreover, if visible, the method can also estimate the position of them.

Finally, apart from a reflector estimation, the acoustic images introduced here can eventually be used in other applications. Although in this work the obtained acoustic images are used to infer only on the geometric information, they capture both geometric and radiometric (e.g. radiation pattern and reflection coefficients) properties of the environment and therefore they have a number of potential applications, mainly applications aimed at extracting some information about the environment or even predicting the acoustic pressure field in generic positions in space.

The work is organized as follows: in Chapter 1, we give to the reader a general framework covering fundamental aspects of acoustics with particular attention to the nature of sound and its propagation in the environment. We describe also the main characteristics of microphones. In Chapter 2 we introduce the problem, make a short review of existing methods, discuss the parametrization adopted and some basic assumptions. In Chapter 3 we define the domain of acoustic rays here referred as the RaySpace, represent the geometric primitives of interest and map the acoustic measurements in the RaySpace. We formalize also the concept of acoustic image and describe how to build it from the acquired signals. Chapter 4 explains the estimation procedures for the localization of sources and reflectors. We take into account both the cases of configurations with single or multiple reflectors. In Chapter 5, we analyze the statistical performance in terms of second order statistics. This performance analysis is useful not only for system design but it will also be used to improve the estimation accuracy. Finally, in Chapter 6 the simulation results validate the performance analysis and experimental results illustrate the applicability of the method in a real world scenario. The conclusions will summarizes the results, makes the final remarks and suggest the future work directions.

Chapter 2

Fundamentals of Acoustics

This chapter aims to give the reader a general framework covering fundamental aspects of the nature of sound and its propagation in the environment. A last section will be devoted to the characteristics of microphones. The concepts and terminology presented here will be used throughout the work.

1 The physical nature of Sound

1.1 Sound propagation

Sound is a mechanical wave propagating in elastic media, as gases, liquids and solids. Sound is transmitted through gases, plasma, and liquids as longitudinal waves, also called compression waves. Longitudinal sound waves alternate pressure deviations from the equilibrium pressure, causing local regions of compression and rarefaction. They are caused by an object, the sound source, which makes vibration movements and transmits its movement to adjacent particles due to the mechanical properties of the medium. The excited particles are displaced from its original position and elastic forces of the medium tend to restore it. Because of the inertia of the particles, they overshoots the resting position, bringing into play elastic forces in the opositie direction, and so on. These particles in turn trasmit their motion to those closest causing a local pressure variation, [4].

Through solids sound can be transmitted also as transverse waves that are waves of alternating shear stress at right angle to the direction of propagation. However, in this work we consider only the transmission of sound in air as

longitudinal waves.

Sound speed, $c[m/s]$, depends on the medium, and other factors. More dense is the molecular structure, easier it is for the moleculas to transfer sound energy. More precisely, it can be proved that:

$$c = \sqrt{\frac{E_\nu}{\rho}}, \tag{2.1}$$

where E_ν is the volumetric elastic modulus for the medium $[N/m^2]$ and ρ is the density of the medium $[Kg/m^3]$.

In the specific case of air there are useful approximated formulas as:

$$c = 331.4 + 0.6t, \tag{2.2}$$

where t is the temperature $[^\circ C]$, [5].

1.2 Acoustic wave

Sound waves can be represented graphically using a Cartesian graph, showing the time (t) on the horizontal axis, and the acoustic pressure (p) of a fixed point on the ordinate. The acoustic pressure $p(t)$ is defined as the difference between the pressure at one point at time t and the static pressure that would exist in the absence of acoustic phenomenon. There are different types of sound waves, and each is identified by a particular graph pattern, [5].

The simple waves can be represented with sinusoidal graphs (Figure 2.1) and correspond to pure tones perceived as a whistle more or less acute depending on the frequency.

The majority of sound waveshapes depart radically from the simple sine wave and in many cases they have not even a periodic trend. Thanks to the Fourier theorem each wave, also aperiodic, under certain mathematical conditions (always verified for physical signals) can be written as a sum (possibly infinite) of simple harmonic waves. Given a time signal $x(t)$, we define the Fourier transform of $x(t)$ as a frequency signal $X(f)$ defines as:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt, \tag{2.3}$$

and the amplitude and phase spectrums of the acoustic wave as:

$$A(f) = |X(f)|, \qquad Q(f) = \angle(X(f)). \tag{2.4}$$

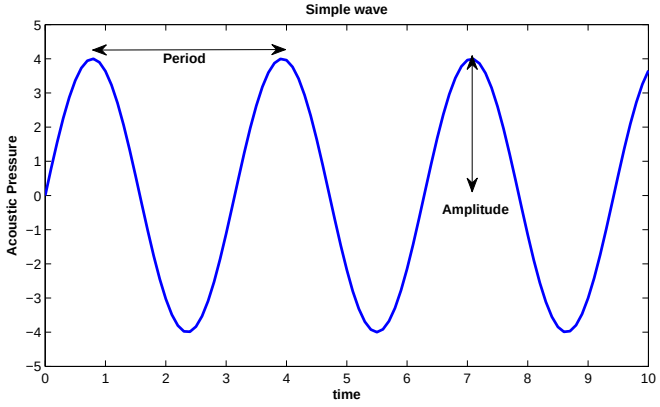


Figure 2.1: Simple sinusoidal wave (pure tone).

Amplitude and phase spectrums contain all the information about the original signal, but in a different form. The Amplitude spectrum allows to show the magnitude associate to all the constituent frequencies of the acoustic signal.

In the particular case of a periodic signal, its Fourier transform is a discrete set of values, which in this case is called the discrete spectrum. The lowest frequency is called the fundamental harmonic and it is the one that has more weight in the final reconstruction of the signal. Other frequencies are multiples of the fundamental and are called secondary harmonics. In this case, the corresponding inverse formula of synthesis is the Fourier series of the original periodic signal, [4] [6].

1.3 Density of energy & Sound Intensity

In this section we are going to introduce other physical quantities useful for the study of the energy characteristics of an acoustic field. The root-mean-square (rms) sound pressure is integral average of the square of the sound signal pressure over a given duration:

$$p_{rms} = \sqrt{\frac{1}{T}\int_0^T p(t)^2 dt}, \tag{2.5}$$

where T is the time of integration for the signal. The rms pressure is most often used to characterize a sound wave because it is directly related to the energy carried by the sound wave, which is called the intensity.

The acoustic intensity of a sound wave is the average amount of energy transmitted per unit time through a unit area in a specified direction. The units of intensity are $[W/m^2]$. The intensity is obtained as the product of the sound pressure and the particle velocity, $\vec{v}$:

$$\vec{I} = p\vec{v}. \tag{2.6}$$

The direction of the intensity is the average direction in which the energy is flowing. For a plane progressive wave, the modulus of acoustic intensity is:

$$I = \frac{p_{rms}^2}{Z}, \tag{2.7}$$

where $Z = \rho c$ is the characteristic acoustic impedance (ρ is the density of the medium and c is the speed of sound). Sound intensity level or acoustic intensity level is a logarithmic measure of the sound intensity, in comparison to a reference level and it is measured in decibel(dB):

$$L_I = 10log_{10}(\frac{I_1}{I_0}). \tag{2.8}$$

The standard reference sound intensity is $I_0 = 10^{-12}W/m^2$.

The sound energy density D describes the time medium value of the sound energy per volume unit; it gives information about the sound energy which is at a defined place in the room, [5]. The sound energy density for an even-proceeding sound wave is:

$$D = \frac{I}{c}. \tag{2.9}$$

2 Acoustic Fields

2.1 The Free Field

We define as free field a sound field generated in a homogeneous and isotropic medium. Although many practical problems of acoustics are associated with structures, such as buildings and rooms, the analysis of sound in the free field is useful because it allows us to understand the nature of sound waves in this undisturbed state. Then, these basic characteristics can be adapted to more complex problems.

We consider a punctual source of sound in a free field. It is easy to understand that the sound intensity decreases as the square of the radius of the distance from the point-source (as the area of any small section on the surface of the sphere increases as the square of the radius, Fig. 2.2).

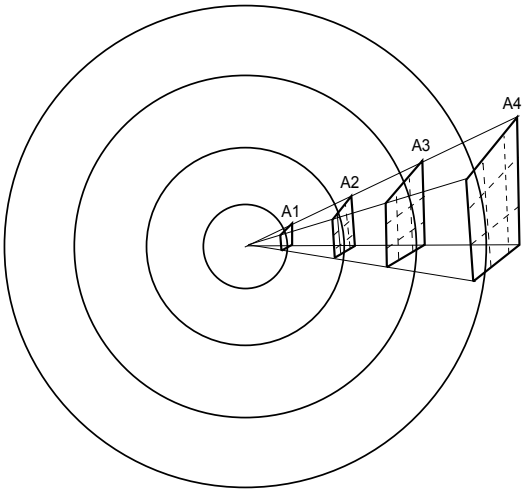


Figure 2.2: The same sound energy is distributed over spherical surfaces of increasing area.

More specifically:

$$I = \frac{W}{4\pi r^2}, \tag{2.10}$$

where I is the modulus of the intensity of sound per unit area, W is the power of source and r is the distance from the source.

Because power per unit area is proportional to the square of acoustic pressure, for the pressure it can be proven that an inverse distance law esists:

$$P = \frac{k}{r}, \tag{2.11}$$

where P is the acoustic pressure and k is an appropriate constant, [4]. For every doubling of the distance from the sound source, modulus of intensity per unit area is quartered and pressure si halved.

2.2 Sound & Structures

When an acoustic wave encounters a wall along its path, the incident acoustic energy (E_i) partially goes back to the room (E_r), partially is absorbed by the wall (E_a) and partially is transmitted outside the room (E_t).

We define the reflector coefficient $r = \frac{E_r}{E_i}$, the absortion coefficient $a = \frac{E_a}{E_i}$ and the transmission coefficient $t = \frac{E_t}{E_i}$. It is easy to understand that the following relation will be true:

$$a + t + r = 1. \tag{2.12}$$

In the room acoustic usually is used a coefficient of apparent absorption $\alpha = a + t$, [5].

The absorption characteristics and the absorption coefficient of a wall vary with the angle of impiges upon the material and so it should be indicated as a_θ where θ is the angle of incidence. In an established diffusive sound field in a room, sound is travelling in all possible directions. In many cases we need an absorption coefficient that is averaged over all possible angle and that is called random incidence absorption coefficient. It is usually referred as the absorption coefficient of the material a.

The absorption coefficient of a material varies with the frequency. Coefficients are typically published at the standard frequencies of 125, 250, 500, 1000 and 4000 Hz. In some cases, the absorption coefficient of a material can be given as a single number called noise reduction coefficient (NRC), that is the average of the coefficiets for $500, 1000$ and 2000 Hz, [4].

The energy that comes back to the room consists of the reflected part, of the diffused part and of the part assorbed that returns to the room due the elastic properties of the wall. Reflection depends partly on the size of the reflecting object. Sound is reflected from objects that are large compared to the wavelength of the impinging sound. Sound reflections follow the same rule of light: the direction of propagation of incident and reflected wave lie on the same plane and the reflection angle, calculated from the normal to the plane of incidence, is equal to the angle of incidence.

The wave reflected from a surface propagates as if it was originated from a virtual source (Fig. 2.3) located in a symmetrical position relative to the surface with respect to the real source. That virtual source is called image source. When sound strikes more than one surface, multiple reflections will be created, [4].

If the surface on which the wave impacts is not perfectly smooth but presentes small substructures, it could be a not specular reflection (Fig. 2.4). The reflection will continue to be mirrored in the case where the wavelength

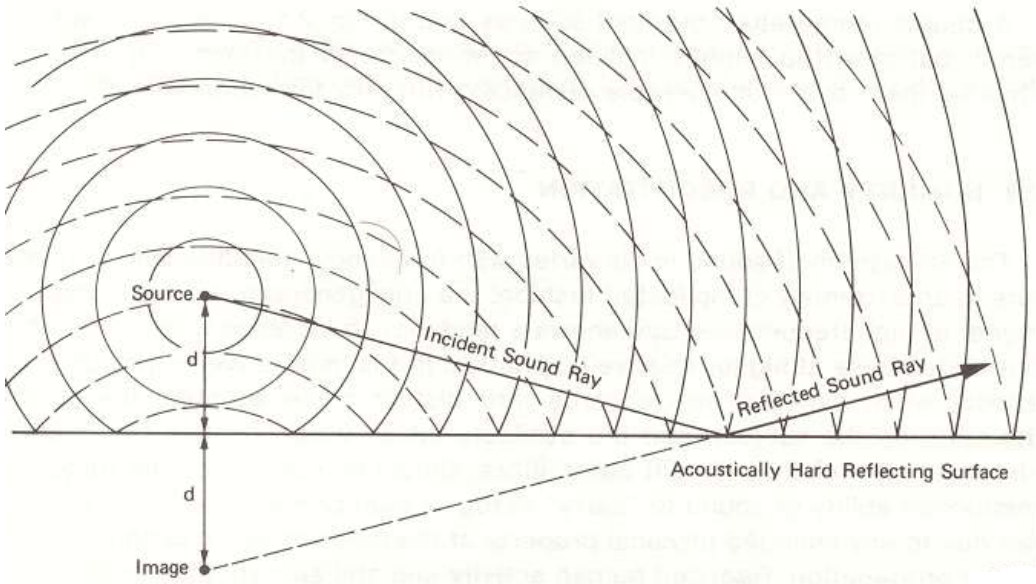


Figure 2.3: Reflections from a plane surface.

of the incident sound is large compared to the characteristic size of surface irregularities. If the wavelength is comparable to the size of the substructures present in the wall, the sound energy is spread evenly in all directions of the environment. This process is known as acoustic diffusion. A perfectly diffusive sound space is one that has the same acoustic properties anywhere in the space. If the wavelength is much smaller, each single surface of the wall will determine independent specular reflections, [4].

Another important phenomenon concerning the interaction between sound waves and structures during the propagation is the diffraction. When a wave encounters an obstacle in its path, it does not fully follow the laws of the optical geometry but it can bend and also the space beyond the line of sight can be affected by it. Huygens formulated a principle that is the basis of the mathematical analysis of diffraction: every point of a wave front may be considered as the source of secondary wavelets that spread out in all directions with a speed equal to the speed of propagation of the waves. According to the principle of Huygens, every point on the wavefront of sound that has passed through an aperture or passed a diffracting edge is considered a point source radiating energy back into the shadow zone. The sound energy at any point in the shadow zone can be mathematically obtained by summing the

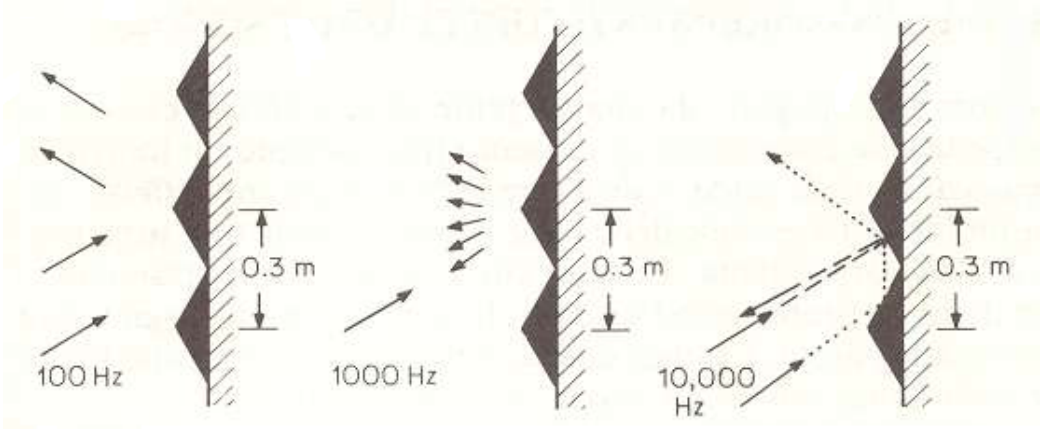


Figure 2.4: Interaction of an acoustic wave with a wall as a function of wavelength ($\lambda_1 = 3.4m$ $\lambda_2 = 0.34m$ $\lambda_3 = 0.034m$).

contributions of all of these point sources on the wavefronts. Low-frequency waves diffract (bend) more than high-frequency waves. That is why diffraction is less noticeable for light than it is for sound because of the relatively short wavelength of light. In general, if the dimensions of the obstacle are comparable to the wavelength of the incident wave, the acoustic perturbation is able to go around it with shadow phenomenon negligible. If the size of the obstacle is greater than the wavelength of the incident wave, occurring phenomena of partial acoustic shadowing, [5].

3 Reverberation

3.1 Early & Late Reflections

Reverberation is defined as the combined effect of multiple sound reflections within a room. The reverberation characteristics of a room are affected by several factors: the shape and size of the room, the materials of which the room is constructed, and the objects present in the room. The materials of the walls are especially important since they determine how much sound is absorbed and how much is reflected, [7].
The impulse response of a reverberating room from a source S to a receiver R can be modeled by splitting the problem into three distinct phases (Fig. 2.5):

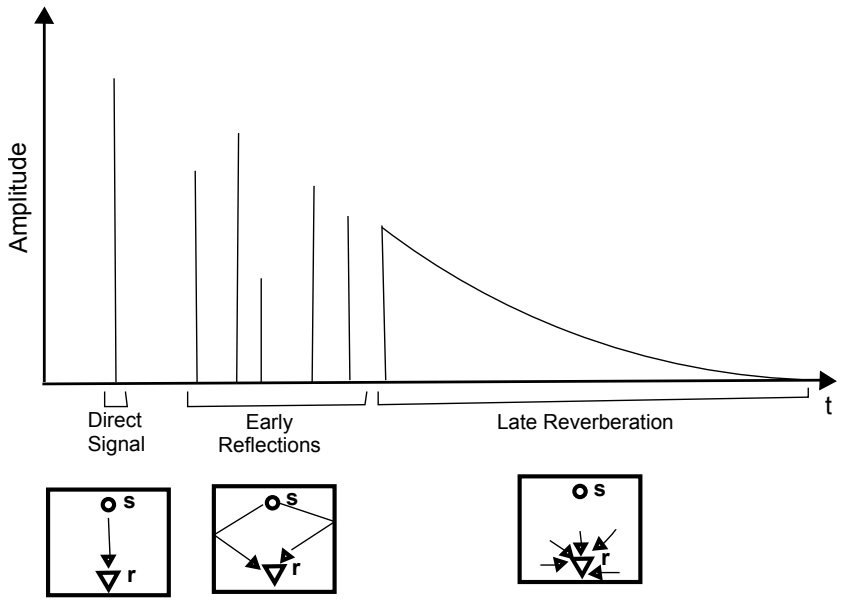


Figure 2.5: Impulsive response of a reverberating environment.

1. The direct signal from the source arrives at the receiver with a delay in time proportional to the distance.

2. The early reflections with the environment occur. This is the most important phase of the impulse response for the perception of the environment in which we are. The early echo pattern depends on the positions and directivities of the source and receiver and they can be described by the image source method. Each image source visible to the receiver contributes a delayed impulse, whose time delay is equal to the distance between the source and the listener divided by the speed of sound. The echo amplitude is inversely proportional to the distance travelled, to account for spherical expansion of the sound, and proportional to the product of the reflection coefficients of the surfaces encountered. Due to a frequency-dependent absorption of incident ray on the wall, frequency-dependent losses should be considered.

3. The late reverberation are no longer distinguishable as separate pulses as a result of the large number of reflections in the room. The energy losses due to dissipation during the wave propagation and energy absorption of the walls cause an exponential decay in the amplitude of

late reverberation. During late reverberation, the dependency on the positions of the source and receiver decreases, [7] [4].

This model describes qualitatively as the sound propagates and decays in a reverberant environment. We want to pose the problem of how to describe quantitatively the characteristics of a reverberant room. For this purpose two different approaches are presented in the following sections.

3.2 Physical approach

The physical approach seeks to describe exactly the propagation of sound from the source to the listener for a given room. From a signal processing standpoint, it is convenient to think of a room with sources and receivers as a system with inputs and outputs, where the input and output signal amplitudes correspond to acoustic variables at points in the room. For example, consider a system with one input associated with a spherical sound source, and one outputs associated with the acoustical pressures at an acquisition device. If the room can be considered a linear, time-invariant (LTI) system (movements in the room are neglected), a transfer function completely describes the transformation of sound pressure from the source to the receiver:

$$y(t) = \int_0^\infty h(\tau)x(t-\tau)d\tau, \tag{2.13}$$

where $h(t)$ is the system impulse response (for the given geometry and source and receiver positions), $x(t)$ is the source of sound and $y(t)$ is the resulting signal. This concept is easily generalized to the case of multiple sources and multiple listeners. By measuring the impulse response of an existing room for the given set of sources and receivers, and then rendering the reverberation by convolution it is possible to obtain the exact propagation of the sound in the environment.

When these measures are not available, we can attempt to predict its impulse response based on purelly physical simulation. This requires detailed knowledge of the geometry of the room, properties of all surfaces in the room, and the positions and directivities of the sources and receivers. Given this prior information, it is possible to apply the laws of wave propagation and interaction with surfaces to predict how the sound will propagate in the space.

The advantage of this approach is that it offers a direct relation between the physical specifications of the room and the resulting reverberation. However, this approach is computationally expensive and stricly tied to the source and receiver locations used for, [7].

3.3 Perceptual approach

The perceptual approach seeks to describe only the perceptually salient characteristics of reverberation. For this purpose we introduce the global descriptors of the reverb. They do not even allow to build a direct input-output relationship as the physical approach to the problem but they give an overall description of the room that is often what it takes for all practical purposes. These descriptors are also less tied to a single experiment (for given positions of sources and reflectors) and give information on the characteristics of the room itself.

We list now the main global descriptors used to characterize a reverberant room:

- EDC is the energy decay curve and it can be obtained by integrating the impulse response $h(t)$ of the room as follows:

$$EDC(t) = \int_t^\infty h^2(\tau)d\tau. \tag{2.14}$$

  The integral computes the energy remaining in the impulse response after time t.

- T_{60} is the time that is required in order to have a decrease of $60dB$ in the energy decay curve:

$$T_{60} = \{t : EDC(t) = EDC(0) - 60dB\}. \tag{2.15}$$

  It is usually taken as the duration time of the reverberation. The reverberant time is directly proportional to the volume of the room and inversely proportional to the amount of absorptions.

- EDR is the energy decay relief, a time-frequency representation of the energy decay:

$$EDR(t_n, f_k) = \sum_{m=n}^{M} |H(m,k)|^2, \tag{2.16}$$

where $H(m,k)$ is the k-th bin of the $STFT$ at time m, and M is the number of temporal frame. It computes the energy remaining in the impulse response after time t_n and in the frequency band centered on f_k. $EDR(0,f)$ gives the power gain as a function of frequency and it is called frequency response envelope, $G(f)$.

- Δf_{max} is the average separation in Hz of resonant modes in the room spectrum. It is approximated as:

$$\Delta f_{max} = \frac{4}{T_{60}}. \tag{2.17}$$

This model is justified for frequencies higher than:

$$f_g = 2000\sqrt{\frac{T_{60}}{V}}, \tag{2.18}$$

where V is the volume of the room.

- N_t is the number of echoes that will occur before time t and it is equal to the number of image sources enclosed by a sphere with diameter ct centered at the listener. The number of image sources enclosed by the sphere can be estimated by dividing the volume of the sphere with the volume of the room:

$$N_t = \frac{4\pi(ct)^3}{3V}. \tag{2.19}$$

From this equation it is clear that the density of echoes grow quadratically with time.

- C is the clarify index and it is one acoustical measure of the direct to reverberant ratio

$$C = 10\log_{10}\left\{\frac{\int_0^{80ms} h^2(t)dt}{\int_{80ms}^{\infty} h^2(t)dt}\right\}. \tag{2.20}$$

This is essentially an early to late energy ratio, which is correlated to the intelligibility of signals in reverberant environments. The time $80ms$ is the reference integration time of human ears, [7].

4 Microphones

4.1 Pressure and Gradient pressure mics

A microphone is an acoustic-to-electric transducer or sensor that converts sound into an electrical signal. Different types of microphones have different ways of converting energy but they all share a mobile component, the diaphragm. This is a thin piece of material which is displaced when a pressure wave occurs. Depending on how the diaphragm is excited we can distinguish two types of microphones: the pressure microphone and the pressure gradient microphone.

The pressure microphones measure the instantaneous pressure produced by a sound wave acting on a diaphragm at a point of the sound field in which it is placed. In Figure 2.6, we show a section view and a front view of a pressure microphone. Only one side (front) of the microphone diaphragm is exposed to the sound field and the other (rear) side is sealed off by a soundproof case. So the diaphragm will be vibrated by changes in sound pressure only at the front side. A small capillary tube connects the interior air mass to the outside, providing a slow leakage path so that static atmospheric pressure will equalize itself on both sides of the diaphragm under all pressure conditions. Very small holes in the backplate are usually evenly distributed on a uniform grid. During the actual back and forth motion of the diaphragm, the air captured in the holes provides damping of the diaphragm's motion at its principal resonance, which is normally in the range of $8 - 12kHz$.

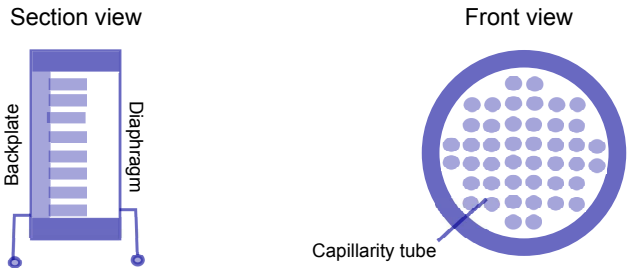


Figure 2.6: Section view and front view of a pressure microphone.

A perfect pressure microphone responds identically to a change in pressure originating in any direction, and therefore arriving at the diaphragm from any angle of incidence. This is why they are also called omnidirectional

microphones.

The pressure gradient microphone senses sound pressure at two very closely spaced points corresponding to the front and the rear side of the diaphragm. In fact there isn't a soundproof case but the diaphragm is in contact with the environment at both ends. Conseguently the motion is driven by the difference, or gradient, between the two pressures. The gradient pressure microphone response depends on the propagation angle of the incident wave.

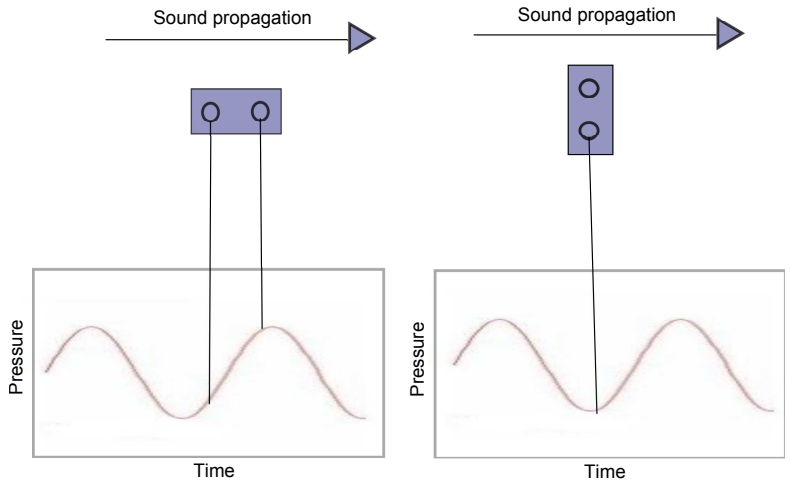


Figure 2.7: Pressure gradient when placed longitudinaly and trasversaly in a progressive plane wave.

In Figure 2.7, we show two opposite situation of trasversaly and longitudinaly position and the pressure acquired in two different points. If the two points are positioned trasversal the wave propagation, there will be no pressure gradient between them and thus no signal at the microphone's output. In contrast, the response is maximum in the longitudinal case. More precisely, the directional sensitivity response equation in polar coordinates is:

$$\rho = \cos(\theta), \tag{2.21}$$

where ρ represent the gain in magnitude of the response and θ is the polar angle.

In Figure 2.8, we show the basic figure-8 response of the gradient microphone in decibel. It is important to notice that the response is maximum at $0°$ and $180°$ but the polarity of the signal is negative in the back hemisphere

relative to the front hemisphere (the diaphragm displacement are in the opposite direction). The frequency range over which the desired polar response is maintained depends on the size of the microphone and the effective distance from the front of the diaphragm to the back. The first null in response take place when the received frequency has a wavelength that is equal to the distance between the two valued points. Because they are sensitive in two directions, pressure gradient microphones are also called bidirectional microphones, [8].

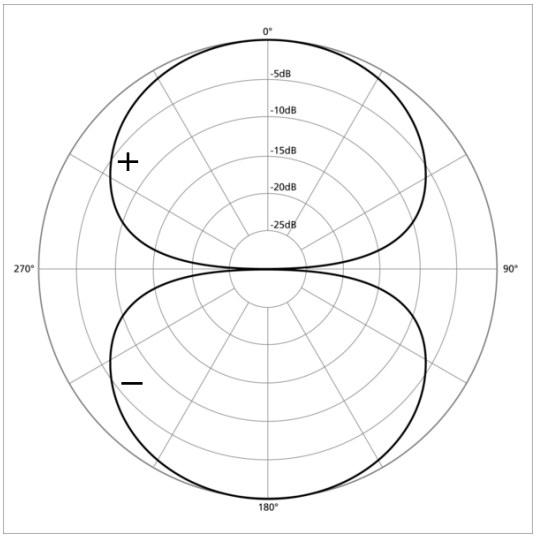


Figure 2.8: Directional response in polar coordinates of the gradient microphone, response level in decibel.

4.2 First order directional microphones

The great majoriy of directional microphones used today are members of the first-order cardioid family. The term first order refers to the polar response equation and its inclusion of a cosine term to the first power. These microphones derive their directional patterns from the combination of a pressure microphone and a gradient microphone.

The earliest directional microphones actually combined separate pressure and gradient elements in a single housing, trasducted both signals and combined their outputs electrically to achieve the desired pattern. Today, most directional microphones with a dedicated pattern have a single diaphragm

and impose a calculated front to back time delay in the paths of sound to achieve the desired pattern.
In Figure 2.9, we summarize the mechanical principle for the directional microphone obtains with a single diaphragm.

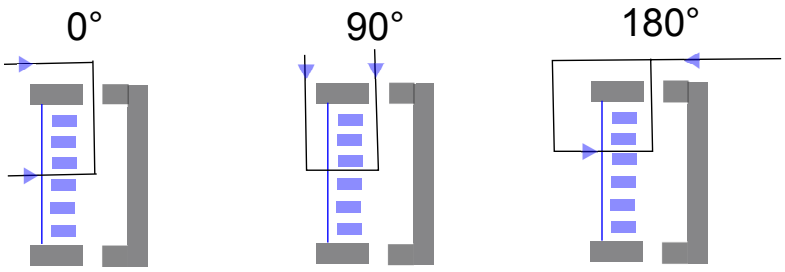


Figure 2.9: Mechanical views of a single diaphragm cardioid microphone.

The general form of the polar equation for the first order directional microphones is:

$$\rho = A + Bcos(\theta), \tag{2.22}$$

where $A + B = 1$. Figure 2.10 shows the graphs and the coefficients of the four main first order cardioid patterns in logarithmic scale.

For some specific applications (e.g film,sports with high ambient noise) it is necessary to use a microphone with directional properties exceeding those of the first-order family. High directionality microphones generally fall into three categories:

1. Interference-type microphones. These designs achieve high direcionality by providing progressive wave interference of high frequency sound arriving off-axis, thus favorite sound arriving on-axis.

2. Focusing of sound by means of reflectors and acoustical lenses. These designs are analogous to optical methods.

3. Second and higher-order designs. These microphones make use of multiple gradient element to produce high directionality.

For the high order design category, a microphone of order p will have the following general form of the polar equation:

$$\rho = A + Bcos^{p}(\theta). \tag{2.23}$$

4.3 Directional properties

Offaxis response can seriously alter the sound of a microphone in real-world use, since in a room there are reflections of the direct sound picked up off axis and combined with the on-axis signal. The combination can radically alter the transduced sound quality through frequency-dependent reinforcements and cancellations. For directional microphones, an important characteristic is the ability of the transducer to select sounds coming from the on-axis direction and reject those coming from other directions. In this section we introduce some microphone characteristics related to the polar response pattern and that describe the ability of the microphone to focus on the desired direct sound and reject ambient sound:

- Acceptable Angle (AA) for a microphone is defined as the frontal angle where the sensitivity is within $3dB$ of the on-axis response. In the case of an omnidirectional pattern, all angles of incidence have a sensitivity of $0dB$ relative to the on-axis response of the microphone. Conseguently, the acceptance angle is $\pm 180°$.

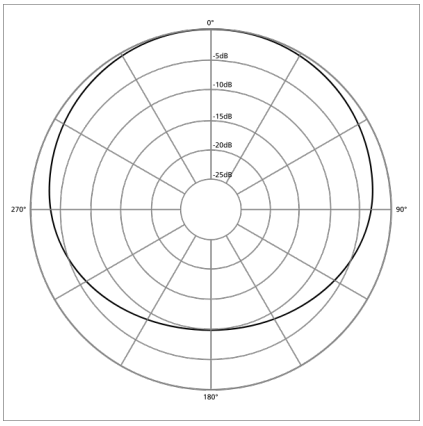
- Random Efficiency Energy (REE) is a measure of the on-axis directivity of the microphone, relative to its response to sounds originating from all directions. An $REE = 0.333$, for example, indicates that the microphone will respond to reverberant acoustical power arriving from all directions with one-third the sensitivity of the same acoustical power arriving along the major axis of the microphone.

- Distance Factor (DSF) is a measure of the "reach" of the microphone in a reverberant environment, relative to an omnidirectional microphone. For example, a microphone with a distance factor of 2 can be placed at twice the distance from a sound source in a reverberant environment, relative to the position of an omnidirectional microphone, and exhibit the same ratio of direct-to-reverberant sound pickup as the omnidirectional.

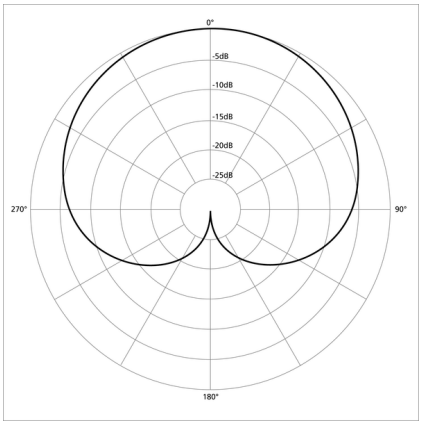
In the table 2.1 we summarize the properties of the first order cardioid microphones, [8].

	Polar Eq	AA	Output 180°	REE	DSF
Omnidirectional	1	$\pm 180°$	$0dB$	1	1
Bidirectional	$\cos(\theta)$	$\pm 45°$	$0dB$	.333	1.7
Subcardioid	$.7 + .3\cos(\theta)$	$\pm 90°$	$-8dB$	.55	1.3
Cardioid	$.5 + .5\cos(\theta)$	$\pm 60.5°$	$-\infty$	.333	1.7
SuperCardioid	$.37 + .63\cos(\theta)$	$\pm 57.5°$	$-11.7dB$	.268	1.9
HyperCardioid	$.25 + .75\cos(\theta)$	$\pm 52.5°$	$-6dB$	.25	2

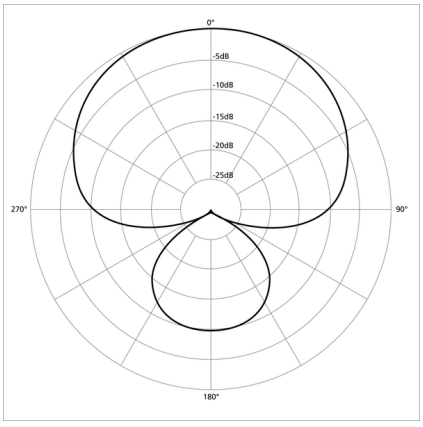
Table 2.1: Characteristics of the family of first-order microphones.



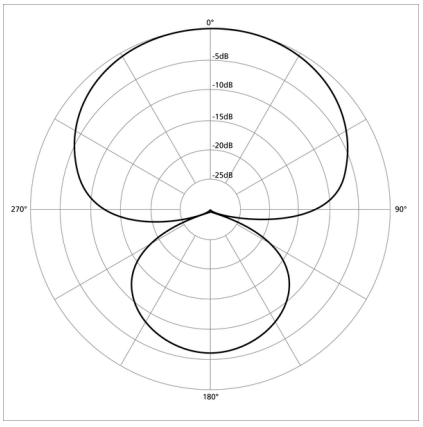
(a) Subcardioid($A = 0.7, B = 0.3$)



(b) Cardioid($A = 0.5, B = 0.5$)



(c) Supercardioid($A = 0.37, B = 0.63$)



(d) Hypercardioid($A = 0.25, B = 0.75$)

Figure 2.10: Logarithmic polar graphs for the first order cardioid family.

Chapter 3

Problem formulation and related work

In this chapter we will give an introduction to the previous works on the problem of finding reflective surfaces in the environment. We will present the hypothesis and the basic ideas of these methods. Then, in the second section, we will formulate the problem as taken into account in the present work and introduce the parametrization that will be used after. The second section will also emphasize on the relaxed assumptions made in this work compared to previous works.

1 Previous 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 (usually only the first most significant reflections are considered) but use different equipment and data. They differ in the number of microphones, the number and knowledge of acoustic source positions, the knowledge of emitted probing signal and the synchronization with the microphones. Mostly the methods localize the reflector matching the estimated room impulse responses ($RIRs$) with template reflections or building constraints from measured 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 [9], it is presented a solution which uses only a microphone and an omnidirectional loudspeaker. The loudspaker rotates on a circular pattern in a continuous fashion and emits a controlled noise and the microphone is located at the center of the circle. The rotation of the loudspeaker induces a time-dependent impulse response between the microphone and the loudspeaker and makes it possible to discern reflections coming from objects located at different positions. In particular, A likelihood map is built by means of a template matching between the signal acquired at the microphone and a template signal obtained by simulating the propagation to all the potential obstacles locations. The reflector position is found in correspondence of the maximum of the likelihood map.

In [10], the necessity for a priori knowledge of the source signal is removed. The approach is based on the inverse mapping of the multi-path propagation problem and acoustic source localization.

In [11], it is presented another method for reconstructing the 2D geometry of the surrounding environment using a single microphone standing in a known position and a loudspeaker rotating around the microphone. The microphone receives, along with the direct signal, delayed and dimmed replicas of the signal associated to wall reflections. A knowledge of the probing signal allows the 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 can be represented as a line of tangency to all the ellipses. So the reflector is estimated by a minimization of a cost function obtained as a combination of quadratic constraints that represent the tangent to such elliptical sufaces. In case of multiple reflectors the $TOAs$ are labelled by means of a generalized Hough transform.

This approach is modified in [12] to take account of not controlled emission and unknown source position. The method is based on estimates of at least three acoustic impulse responses ($AIRs$) between stationary microphones located at arbitrary but known relative positions and a sound source located at an unknown position. A Two-step approach is uses. At first the

authors estimate the location of the source relative to the position of a reference microphone and the time differences of arrival ($TDOAs$) associated with direct paths. Secondly, they use the estimated position for the estimation of the $TOAs$ related to first-order reflections, to constrain the possible reflector locations as in the previous work. In case of multiple reflectors the related ellipses are grouped iteratively.

The same approach is further extended in subsequent work. The previous methods for estimation of the line reflector rely on solving a non-linear and non-convex optimization problem that can converge to non-optimal solutions in certain practical environments. In [13], authors address some of these problems by deriving a closed-form solution for the single-reflector case, which yields a more robust solution when compared to the other approach. The article proposes also a second-stage correction in order to improve robustness of the method in the presence of errors in the estimated $TOAs$. This consists in find meaningful coordinate points, that are geometrically related to the estimated line reflector from the closed-form solution and the set of ellipses related to the $TOAs$ and estimate the reflector by the points that exhibit greatest collinearity.

In [14], another way to solve the problem of non-optimal solutions for the non-linear and non-convex optimization for [12] is presented. The authors propose an exact minimization procedure that determines the correct global minimum of the cost function while circumventing the problem of local minima. This result is obtained by the reformulatation of the probelm as the constrained minimization of a second-order polynomial, which admits an exact solution. In this paper is also proposed a methodology for error propagation analysis, which aims to characterize the error that the reflector localization is affected by using some prior information on the error on TOA measurements.

In [15], the approach in [12] is generalized for localization of three-dimensional surfaces by planes parametrization in the geometric space. Besides extending the approach to a $3D$ geometry, this paper provides a generalization of the Hough transform to arbitrary configurations of source and receivers.

In [16], the authors use a different approach using the estimation of the Directions of Arrival ($DOAs$) for different positions of the acoustic source instead of $TOAs$. That is an interesting evolution beacause the measure-

ment of the Time Of Arrival of the reflected path implicitly requires the synchronization between source and microphones, which is not viable in many applications. Once the location of the source and the DOA of the reflected path are estimated, the line that parameterizes the position of the obstacle is constrained to be tangential to a parabola having the focus in the source and directrix the measured DOA. A parabola is easily described by its matrix quadratic form, as for ellipses and, as in [12], a cost function is defined as combination of multiple constraints. Once again multiple constraints are obtained for multiple positions of the source and the reflector is estimated with a minimization of the cost function subject to these constraints.

In [17], is presented a method for the localization of major reflectors in a room that use robust beamforming techniques based solely on the recorded microphone signals of a circular array. The main advantage of this method is that it does not involve measuring room impulse responses and so, it can generally be applied for any source signals which sufficiently excite all room modes. The method consists of the three-step procedure: first, the ($DOAs$), corresponding to all sources and reflections are determined, then the signals originating from these $DOAs$ are extracted, and finally, the time differences of arrival($TDOAs$) are estimated from crosscorrelation analysis of the extracted direct sound and its reflections, from which the distances from the circular array to the reflectors are inferred. Convex-optimized beamformers are used for both localization and extraction of early reflections in order to control the robustness of the beamformersi compared to the signal to noise ratio. Similar procedure is performed in $3D$ environments in [18] using a spherical microphone array and an adaptation of the Convex-optimized beamformer technique.

It is also worth mentioning the work in [19] where consider the problem of estimating room geometry from a single acoustic room impulse response (RIR), a $1-D$ function of time. The uniqueness of the mapping between the geometry of a planar polygonal room and a single RIR is verificated and it is presented an algorithm that performs the "blindfolded" room estimation. The results are used to construct an algorithm for localization in a known room using only a single RIR.

Finally in [20] authors propose a method for reflector localization in simple 3D environments using a fixed compact array of M microphones with an

integrated and synchronized loudspeaker that probes the environment with a known test signal. To identify strong reflectors, they propose to use a regularized least-squares procedure and fit known reflection templates to measured $RIRs$. It produces a sparse set of strong reflections with known DOA and range. These reflections are analyzed and further classified into 1st, 2nd and 3rd-order reflections or clutter, from which the room model can be correctly inferred.

2 Problem formulation and parametrization

In this work we use a linear microphone array and a moving acoustic source with unknown position end emitting signal. The presented work is developed for $2D$ geometry and we consider only specular reflections, assume $2D$ propagation (i.e. perfectly absorbing floors and ceilings). The extension to $3D$ cases can be done with some considerations and modifications, using a planar array.

With reference to Figure 3.1, let us suppose to have a microphone array that captures the soundfield produced by an acoustic source within a $2D$ environment. Our objective is to estimate the position of the reflective surface in the space. We remove the hypothesis of infinite extension of reflective surfaces taking into account the limited visibility of acoustic reflections and, in particular we are not interested in estimating only the line that contains the reflector but also its endpoints (points A and B in Figure 3.1).

The presence of reflectors will cause reflections to appear. According to the Snell's law they can be thought as originated from image sources obtained by mirroring the source with respect to reflectors (see S and S' shown in Figure 3.1). However the visibility of image source is limited in space by the extension of the reflective surface. In fact the rays originated from the image source are bounded to intersect the reflector as shown in the Figure 3.1. The observation of a number of this visibility discontinuities, together with the estimate of the image source positions, will allow us to obtain estimates of A and B.

If we have a moving source that "illuminates" the environment, we can take various acoustic images at different times of the space surroundings. Our objective is the reconstruction of the geometric properties of the room

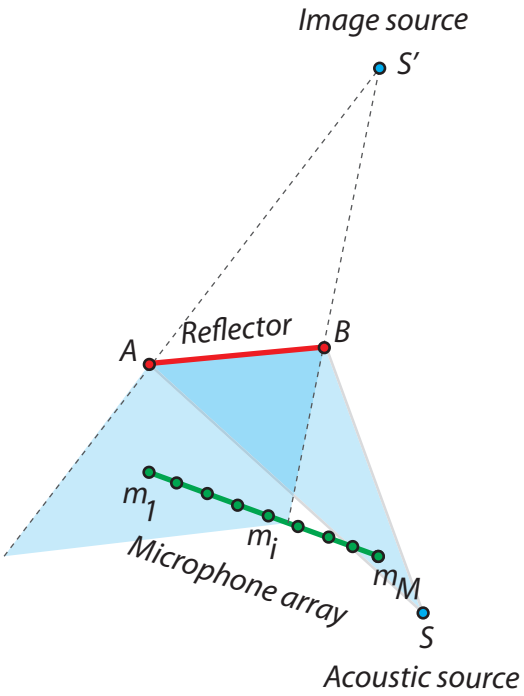


Figure 3.1: Problem setup: a microphone array captures the suondfield produced by an acoustic source within an environment with a single reflective surface.

using these images. It's important to notice that the direct source and its sound emissions does not have to be controlled. The microphone array and loudspeaker are not synchronized. The probing signal and source positions are not necessarily known. We need only that the environment will be excited from different positions in space in order to guarantee acquisition of different "views" of the acoustic scene. That means that, in a real scenario, we can use real and uncontrolled sound sources for the acquisition of the acoustic images and the estimation of the desired geometric properties of the environment.

Reconstruction of the environment from a series of images is a well known problem in computer vision. However, what is visible in optics is not necessarily 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 is not easily extended to acoustics. First we must define the concept of "acoustic image", i.e. represent the acoustic measurements with suitable coordinates.

In optical field, the images are defined as a sampled version of the Plenoptic function, introduced in [21]. The Plenoptic function defines the intensity of the light flow for every possible location, at every possible direction, for every wavelength, and, for dynamic scened, at every time. It is an idealized concept, and one does not expect to completely specify it for a natural scene. Obviously one cannot simultaneously look at a scene from every possible point of view, for every wavelength, at every moment of time. But, by describing the Plenoptic function, one can examine the structure of the information that is potentially available to an observer by visual means. When you want to characterize the function for a real environment, you measure and define a sampled version of Plenoptic function in most important variables of dependencies. Usually several assumptions (e.g. static scenes, grayscale images, reduction of degrees of freedom on camera locations) are made in order to reduce the dimensionality of the problem [22].

Popular parametrizations of the Plenoptic function are Lumigraph [23] and Light Field [24]. The Lumigraph is a subset of the complete Plenoptic function that describes the flow of light at all positions in all directions. It considers only the subset of light leaving a bounded object (or equivalently entering a bounded empty region of space), exploiting the fact that radiance along any ray remains constant (in a trasparent medium as the air). That allows to reduce the domain of interest to four dimensions. The Light Field uses a similar approach in order to map the Plenoptic function in the space of oriented lines.

In order to make the reconstruction of the environment from an acoustic point of view, we think of the acoustic image as a sample of the Plenacoustic function that has the same role of the Plenoptic function in optics. The Plenacoustic function was at first introduced in [25] as a 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 [25], this omnidirectional function can be turned into a directional one using the phase information.

For the purpouse of this work we consider the Plenacoustic function as a function of position and direction. In order to make this extension to a directional function, we will use the space-time processing techniques to estimate the power coming from a given direction of arrival (DOA). That allows to

take sampled version of the directional Plenacoustic function (the magnitude of a set of rays in the environment) using a microphone array. As a consequence we represent geometric primitives of interest (rays, sources, receivers and reflectors) and the acoustic measurements (i.e. the acoustic image) in the space of oriented lines here referred as the RaySpace. It's important to note that the primitives representation in the RaySpace will be an injective map and so the problem to estimate the geometrical properties of the environment in the geometric space is equivalent to search these estimation in the RaySpace. So, our goal will be to extract geometric informations on the environment (i.e the primitives representation into the RaySpace) from the acquired images (i.e the estimated magnitude of the rays).

The domain, the RaySpace, is the same used in [26] for rendering applications. However in [26] the environment representation in the RaySpace were the starting point for the modelling of acoustic propagation in generic environments. Here we would like to turn them into the result of our analysis and thus we need to reverse the problem.

Chapter 4

The Acoustic Images

In this chapter we present the RaySpace parametrization for the acoustic rays. In the first section, we map into the RaySpace a minimal set of foundamental objects that are sufficient to characterize the propagation in the environment. In the second section, we will present the ideal acoustic camera and its real sampled version (the microphone array) that is used in order to capture acoustic rays.

1 Geometric primitives

1.1 Representation of a ray

The acoustic ray can be seen as an oriented line in the geometric space. A line in $\mathbb{R}^2$ is represented by the equation:

$$l_1 x_1 + l_2 x_2 + l_3 = 0. \tag{4.1}$$

We parameterize a ray with the coordinates $[l_1, l_2, l_3]^T$ of the line on which the ray lies. We notice that $k[l_1, l_2, l_3]^T$, with $k \in \mathbb{R}$ and $k \neq 0$, represent the same ray. As a consequence, this parametrization defines a class of equivalence, as it uses scalable - homogeneous - coordinates. However, rays have a travel direction. In order to distinguish rays lying on the same line but with opposite orientations, we limit the range of the scalar k to the positive or negative interval:

$$\mathbf{l}_1 = k[l_1, l_2, l_3]^T, \ k > 0,$$
$$\mathbf{l}_2 = k[l_1, l_2, l_3]^T, \ k < 0. \tag{4.2}$$

A generic point in the $(l1, l2, l3)$ space corresponds to a ray in the geometric space and thus this parametrization is here referred as the RaySpace. The equivalence class inherent in the RaySpace implies that the RaySpace is a projective space $\mathbb{P}^2$ that span the Eucledian space ($\mathbb{R}^3$) by such homogeneous coordinates of lines.

For clarity of visualization, in this work we usually rather than visualizing the whole three dimensonal RaySpace, we depict the primitives in a reduced $2D$ RaySpace, obtained by intersecting the RaySpace with a prescrived plane as shown in Figure 4.1 (b). We notice, however, that in the reduced RaySpace we cannot distinguish rays with the same direction but opposite orientations. It is not a problem for the purposes of this work because we will use a linear microphone array to capture only the rays coming from a part of the space (i.e. having an appropriate orientation).

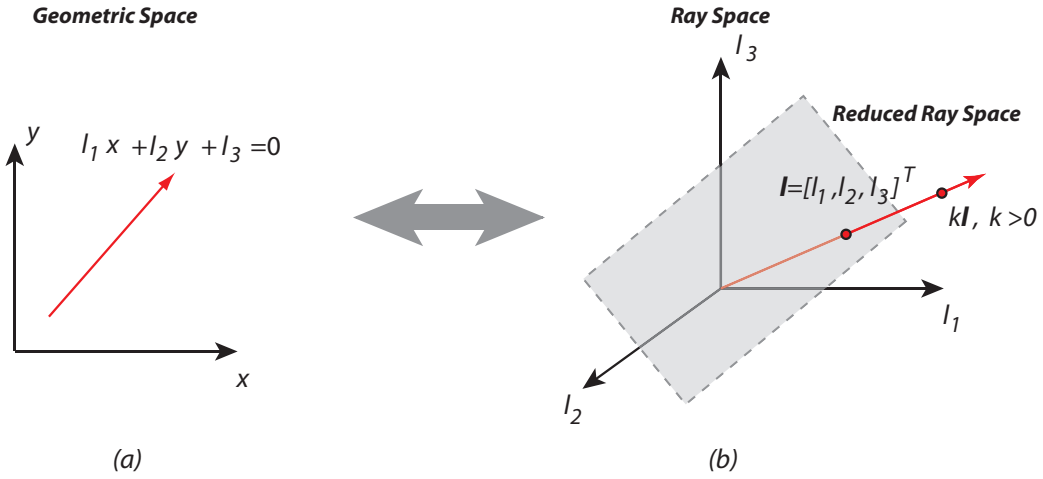


Figure 4.1: A ray in geometric space (a) and RaySpace (b).

1.2 Representation of sources and receivers

Acoustic source and receiver can be seen as points in geometric space. We represent a point $P = (x_1, x_2)$ in the RaySpace by the set of all rays that

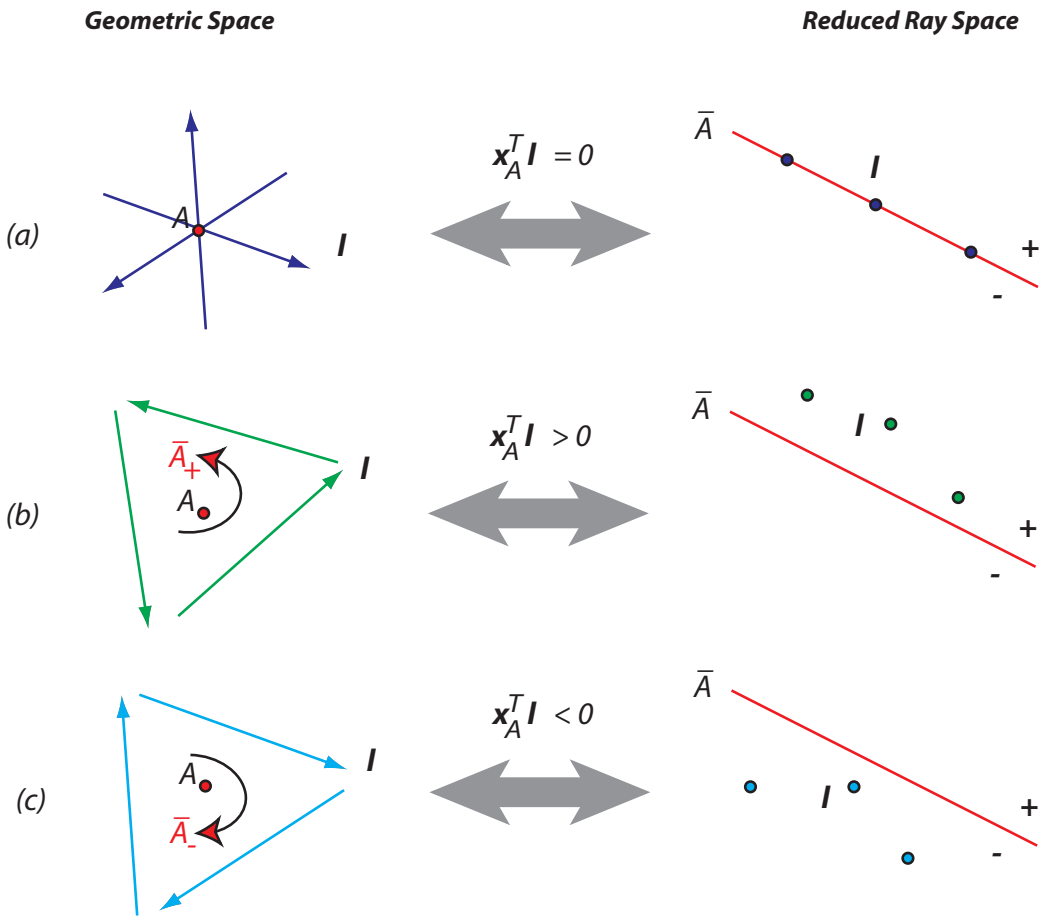


Figure 4.2: A point in geometric and ray space (a); the orientation of rays with respect to a point in the geometric space and their configuration in the ray space (b),(c).

pass through it. From 4.1, a ray $\mathbf{l}$ i passing from a point A with homogeneous coordinates $\mathbf{x}_A = k[x_A, y_A, 1]^T$, $k > 0$ if:

$$\mathbf{x}_A^T \mathbf{l} = 0. \tag{4.3}$$

Using the condition 4.3, all rays passing through the point A are defined as:

$$\overline{A} = \{(l_1, l_2, l_3) \in \mathbb{R}^3 | l_1 x_A + l_2 y_A + l_3 = 0\} = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T \mathbf{l} = 0\}.$$

The previous equation shows that a point (source or receiver) in the geometric space corresponds to a plane passing through the origin in the RaySpace (see Figure 4.2 (a)). As a point in the geometric space is a plane

in the RaySpace, it divides the space into two half-spaces. This allows us to test the orientation of a ray with respect to the point. In particular, all the rays that has the point in the geometric space to their left with respect to the travel direction are given by (Figure 4.2 (b)):

$$\overline{A}_+ = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T \mathbf{l} > 0\}. \tag{4.4}$$

Similarly, all rays that have the point on their right are given by (Figure 4.2 (c)):

$$\overline{A}_- = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T \mathbf{l} < 0\}. \tag{4.5}$$

1.3 Representation of a reflector

In the geometric domain the reflector R is a line segment and it is completely defined by the two endpoints $A = (x_1^A, x_2^A)$ and $B = (x_1^B, x_2^B)$. As for points, we represent the reflector in the RaySpace as the set of all rays that pass through it (through all the intermediate points). In the RaySpace this correspond to the set of all planes representing the infinite intermediate points between A and B.

$$\overline{R} = \overline{A} \cup ... \cup \overline{P_i} \cup ... \cup \overline{B}. \tag{4.6}$$

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. This motivates the definition of two reflectors, one for each face of the line segment. Traditional projective geometry does not account for oriented reflectors, while Oriented Projective Geometry provides the tools required for representing them. With reference to Figure 4.3, the two rays l_1 and l_4 that fall onto the opposite faces of the line segment R in the travel direction have the endpoints A and B on opposite sides: A is on the right for l_1 and on the left for l_4 and therefore $\mathbf{x_A}^T \mathbf{l_1} < 0$ and $\mathbf{x_A}^T \mathbf{l_4} > 0$.

We exploit this inequalities when we represent the two oriented reflector $\overline{R_1}$ and $\overline{R_2}$ corresponding to the not-oriented reflector $\overline{R}$ in the RaySpace:

$$\overline{R_1} = \left\{\mathbf{l} \in \mathbb{P}^2 \left| \mathbf{x_A}^T \mathbf{l} < 0 \right.\right\} \cap \left\{\mathbf{l} \in \mathbb{P}^2 \right| \left. \mathbf{x_B}^T \mathbf{l} > 0 \right\} = \overline{A_-} \cap \overline{B_+}. \tag{4.7}$$

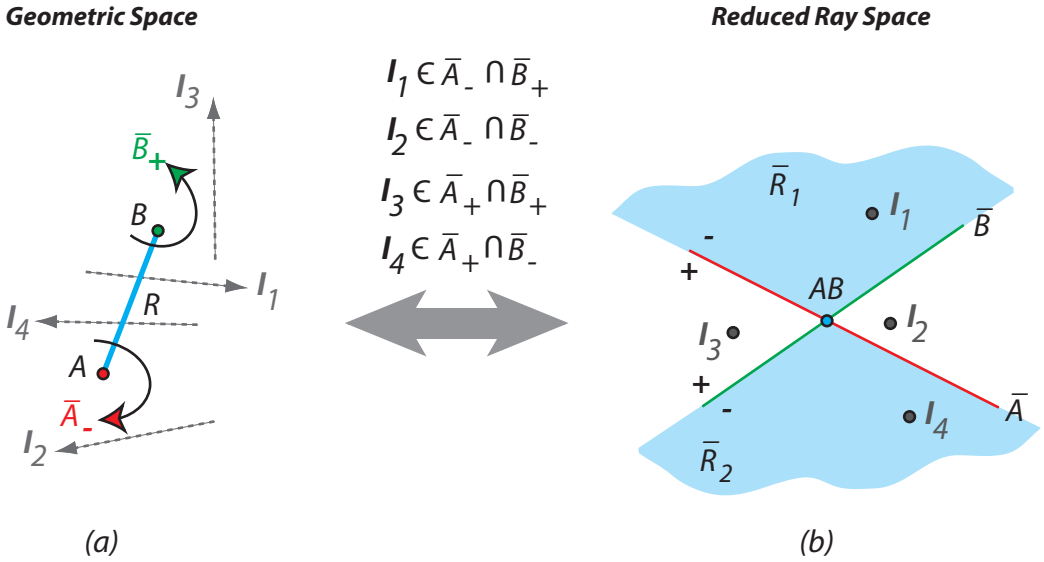


Figure 4.3: Four subspaces defined by the two endpoints in the RaySpace b) and the corresponding rays in the geometric space a).

$$\overline{R_2} = \left\{\mathbf{l} \in \mathbb{P}^2 \,\middle|\, \mathbf{x_A}^T\mathbf{l} > 0\right\} \cap \left\{\mathbf{l} \in \mathbb{P}^2 \,\middle|\, \mathbf{x_B}^T\mathbf{l} < 0\right\} = \overline{A_+} \cap \overline{B_-}. \tag{4.8}$$

Obviously, the rays which have the points on the same side (l_2 and l_3) do not belong to the reflector and they are included in the subspaces $\overline{A_+} \cap \overline{B_+}$ and $\overline{A_-} \cap \overline{B_-}$.

The not-oriented reflector can be expressed in a closed form as the union of the two oriented reflectors that compose it:

$$\overline{R} = \overline{R_1} \cup \overline{R_2} = \left\{\overline{A_-} \cap \overline{B_+}\right\} \cup \left\{\overline{A_+} \cap \overline{B_-}\right\}. \tag{4.9}$$

All the rays originated from an oriented reflector R_i with endpoints A and B form the visibility region from that reflector (the set of all rays that are reflected by R_i). By intersecting this region with the rays that fall onto another oriented reflector R_j with endpoints C and D we obtain the visibility region of R_j from R_i. These rays could be present in case of multiple reflection:

$$\nu(R_i, R_j) = (\overline{A}_{+/-} \cap \overline{B}_{-/+}) \cap (\overline{C}_{+/-} \cap \overline{D}_{-/+}). \tag{4.10}$$

If the environment is composed of more than two reflectors, mutual occlusions could arise. This corresponds to an overlapping of visibility regions in the RaySpace.

2 Acoustic images

2.1 Ideal acoustic camera

Let suppose for a moment to have an ideal acoustic camera capable of capturing the complex amplitude of all acoustic rays that fall onto it. Actually, as the captured rays are not bounded to pass through a single point (camera center), this measuring device represents an ideal Plenacoustic camera. We want to use that camera in order to obtain information about the environment. We suppose to have a linear acoustic camera that is represent as a line segment in the geometric space. As a line segment, it has the same representation in the RaySpace of a reflector (i.e. the set of all rays that intersect the line segment).

In Figure 4.4, we show a possible environment with a reflector R with endpoints A and B and an acoustic camera AC with endpoints C and D. The gray area represents the region of visible by the acoustic camera, which in this case can be interpreted as the set of all rays that can be captured by the camera. The red area represents the rays that intersect both reflector and acoustic camera. It is the region of visibility of the reflector from the acoustic camera, $\nu(R, AC)$, and it contains all the rays that could be reflected by the reflector and acquired by the acoustic camera. In the case of an ideal acoustic camera, we acquire all the intensity of the rays in the red area in the configuration analyzed. Up to now we discussed only the representation of geometric primitives in function of acoustic rays that are generated (source), are collected (receiver) or are reflected (reflector) by them. Each ray that parametrize the geometric primitive will have a corresponding amplitude associated with it.

However, without an acoustic source to "illuminate" the environment, the acoustic camera wont acquire anything except a noise. Suppose now to "light up" the environment with an acoustic source. For simplicity in

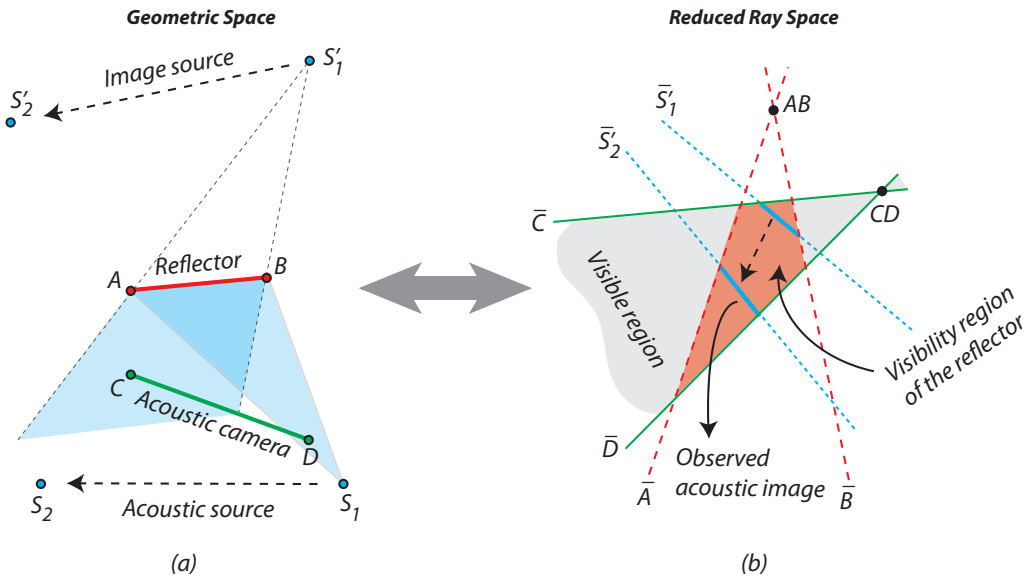


Figure 4.4: The ideal acoustic camera measures the soundfield produced by a source in an environment with a single reflector (a); the ray space representation of acoustic image (b).

Figure 4.4 (a) the source is placed behind the acoustic camera in order to reduce its direct influence on the captured soundfield (we suppose that the ideal acoustic camera sense only the rays that fall onto it by the opposite direction). The source begins to populate the environment with acoustic rays. The presence of reflectors will cause reflections to appear. They can be thought as originated from image source obtained by mirroring the source with respect to reflectors as shown in Figure 4.4 (a). The visibility of image source is limited in space as the rays originated from it that intersect the reflector. Some of this rays will be captured by the acoustic camera, i.e. the rays originated from the image source that belong to the area of visibility of the reflector from the acoustic camera. The intensity of these rays are registered by the camera and these measurements represent a single acoustic image.

Notice that in this ideal case the acoustic camera captures also the diffuse reflections. As a consequence, along with the line segment representing the image source (i.e. the specular reflections), the whole visibility region lights up. Therefore a single Plenacoustic image is enough for reflector estimation. In fact, the visibility regions constitute an equivalent representation of the

reflectors in the RaySpace, and we can reconstruct the environment geometry by looking and examining this ideal Plenacoustic image.

However, unlike optics where most surfaces can be considered Lambertian (all sufaces have a constant radiance or luminance that is independent of the viewing direction), in acoustics the specular reflections dominate the diffusive reflections. Thus the amplitudes of diffuse reflection will have a considerably lower amplitudes and will hardly be observable. Furthermore in real world scenarios we do not have an ideal acoustic camera but we use instead a microphone array. As a consequence we have to take more images illuminating different portions of the reflector's visibility region. This is accomplished moving the source in space (see Figure 4.4 (a) and (b)).

2.2 Microphone array

In real scenarios, an ideal acoustic camera representable by a line segment that is able to acquire punctually the magnitude of all rays that fall into each point of the line and for all the direction of arrival does not exist. Instead of it, we use a linear microphone array that is composed by M microphones m_i with $i = 1, ..., M$.

A real microphone array has several limitation compared to an ideal Plenacoustic camera. The continuos line segment of an ideal acoustic camera is sampled in the microphone array with each node of sampling in the position of a microphone. Consequently, also the acquired acoustic image will be sampled because it is possible to associate a magnitude only to the rays affecting a material point where there is a microphone. So we will have informations about the intensity only of the rays that lie on the planes in the RaySpace corresponding to the microphones positions in geometric space (see Figure 4.5 (a) and (b)).

We define as $P_i(\theta)$ the spatial pseudospectrum of the microphone m_i, the magnitude acquired by the microphone from the direction of arrival θ. We define also as $P_i(\mathbf{l})$ the corresponding magnitude of rays of incidence on the microphone m_i.

A single microphone acquires only the overall intensity of the sound resulting from the contributions of each directions of incidence attenuated according to its own polar pattern. It is not possible to separate the directional contributions (the intensities of the various incident rays) and to obtain $P_i(\theta)$

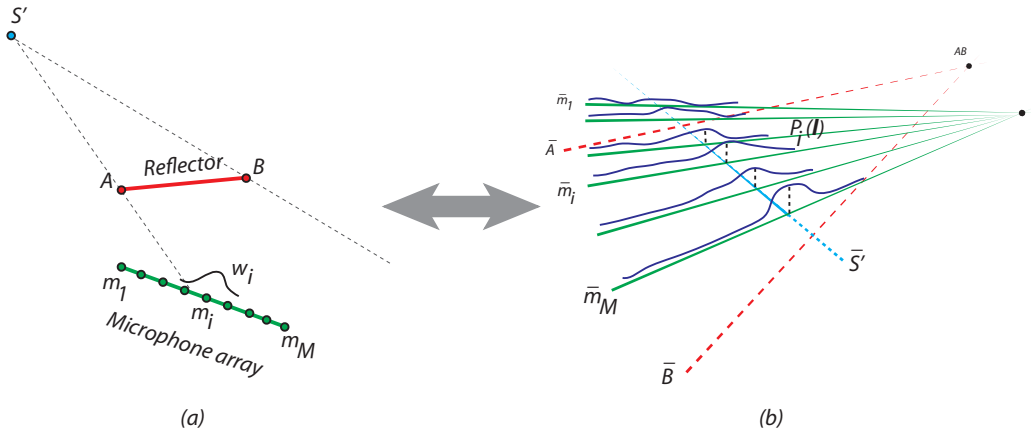


Figure 4.5: The microphone array measures the soundfield produced by a source in an environment with a single reflector (a); the ray space representation of acoustic image (b).

analyzing the acquisition of a single microphone.

So, for the estimation of the acoustic image representing the environment in the RaySpace we have to do the following steps for every microphone m_i, $i = 1, ..., M$:

1. window the contributions of all microphones with the window function w_i centered at m_i;

2. use a beamforming technique to estimate the spatial pseudospectrum $P_i(\theta)$ and associate it to the point m_i;

3. obtain the corresponding $P_i(\mathbf{l})$ mapping the values of $P_i(\theta)$ in the RaySpace.

Any beamforming method that takes into account the correlation between the sources (signals from various paths are attenuated and delayed versions of the same direct source signal) can be applied for this purpose. In simulations and experiments we will use wideband Capon method [27] applied to the linear array. However, if the specific application requires to improve accuracy and resolution, other robust and superdirective methods can be used.

It is important to notice that the width of the window function is a fundamental parameter of the method. To be able to observe a bigger portion of space the microphone array needs to have a big extension, either having one long array or virtually moving a smaller array (e.g. translating along a

rail). On the other hand, in order to obtain estimates of rays going through the point m_i, we have to use microphones close to it with decreasing influence of microphones farther away.

Using a wide window would mean considering phenomena that not affect the considered point m_i. As the image sources are not visible from all points in space we risk also to add only noise increasing the number of contributions without a considerable gain on direction estimation accuracy. What's more important, using a number of microphones with a big extension in space we violate the far-field assumption (the assumption that the DOA is the same for all the microphones in the window) made by the DOA estimation techniques used in this work. It is worthwhile to note that although there are also nearfield techniques, they use information that are not available for our analysis (usually they require the distance of the sound source). We can not use also Music method [28] because it assumes that the (direct and image) sources are uncorrelated and it is not our case.

Once we have the magnitude of sampled DOA, $P_i(\theta)$ from the beamforming method, for all microphones we have to map the values on the RaySpace in order to obtain the $P_i(\mathbf{l})$. Given the coordinates $p_i = [x_i, y_i]^T$ of the microphone m_i, the directions of arrival are mapped into points in the RaySpace using the following non-linear relation:

$$\mathbf{l}(\theta) = [-cos(\theta), sin(\theta), cos(\theta)x_i - sin(\theta)y_i]. \tag{4.11}$$

Then we can associate the intensity of the DOAs for all microphones to the corresponding rays.

Finally we obtain the acoustic image shown in Figure 4.5 (b). It is a sampled and "blurred" version of the image we would obtain using the ideal acoustic camera, as seen in Figure 4.4 (b). In fact, we compute the pseudospectrum for sampled points m_i and for sampled DOA using the beamforming method.

Furthermore, instead of having impulses representing the image source (with amplitudes depending on radiometric properties of the environment), they are convolved with the aperture function of the DOA estimation technique depending on the used window function and the number of microphones involved in the estimation.

Note that the amplitude along the line representing the image source will vary according to the radiation pattern of the source, the polar pattern of the acoustic camera, the reflection coefficients and the travel path distance. The line position into the acoustic image depends also directly by the geometric configuration i.e the position of reflector and of the direct source. This means that in an optic of estimation, if the configuration is unknow we can extract information about it from the image. The beamforming methods return also the phase spectrum related to the environmental response but for the purposes of the work the phase information can be discarded.

In case of wideband sources we can obtain a number of acoustic images for different frequency bands of interest or (if we are not interested in extracting radiometric information) obtain a single image combining images at different frequencies (making a suitable average of the obtained images).

Chapter 5

Estimation algorithms

In this chapter we analyze the sampled and "blurred" version of the acoustic images of the environment acquired by the microphone array in order to extract the estimations of geometrical features. First, we will present algorithms for the localization of a single (direct or image) source and a single reflector. Then, these techniques are adapted for cases of multiple reflectors.

1 Single source localization

First we take into account the localization problem for a single source due the analysis of a single acoustic image. The source to be estimated can be either the direct source that emits a signal in the environment or an image source generated from a single reflector, if the direct source is in a position that makes the direct influence on the captured soundfield negligible (e.g behind a microphone array of directional microphones that point in the opposite direction). For now on, the notation will refer to this second case, but the technique is the same in both the cases.

We start from an image obtained using the microphone array. An example is shown in Figure 5.3 (a). First we have to detect the $N_{S'}$ microphones that see the image source S'. For this purpouse, we calculate the global maximum of the scanned image ($P(\mathbf{l})$) and we consider a microphone m_i if and only if the maximum of the acquired magnitudes for that microphone (i.e. the maximum of $P_i(\mathbf{l})$) is over a threshold defined as a percentage of the global maximum.

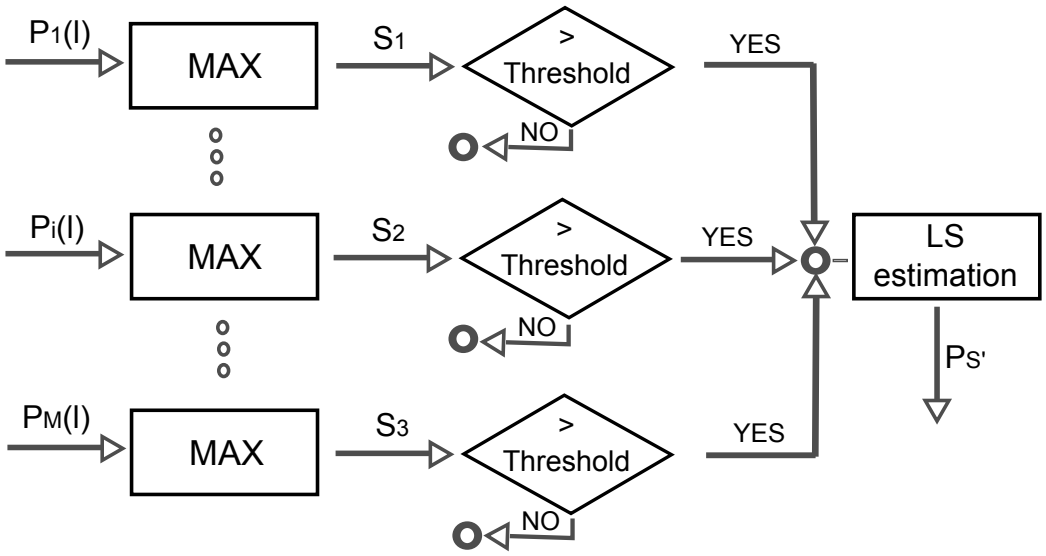


Figure 5.1: Diagram for the algorithm of estimation for a single source localization.

After that we define the rays of acquisition from the image source for the microphone array as:

$$\hat{\mathbf{s}}^i = \arg\max_{\mathbf{l}}(|P_i(\mathbf{l})|), \forall i \in \mathcal{I}_{S'}, \tag{5.1}$$

where $\mathcal{I}_{S'}$ denotes the subset of indices such that m_i with $i \in \mathcal{I}_{S'}$ is "illuminated" by the image source S'. The (homogeneous) points $\mathbf{s}^i = [s_1^i, s_2^i, s_3^i]^T$, $i \in \mathcal{I}_{S'}$, should all stay on the plane $\overline{S'}$ as depicted in Figure 5.3 (a). So, using equation (4.3), we can write:

$$\mathbf{x}_{S'}^T \mathbf{s}^i = 0, \ i \in \mathcal{I}_{S'},$$

where $\mathbf{x}_{S'} = k[x_{S'}, y_{S'}, 1]^T$, $k > 0$, are homogeneous coordinates of S'. Indicating $\mathbf{s}_j = [s_j^{i_1}, s_j^{i_2}, ...]^T$, $i_1, i_2, ... \in \mathcal{I}_{S'}$, for $j = 1, 2, 3$, and $\mathbf{H}_{S'} = [\mathbf{s}_1, \mathbf{s}_2]$, $\mathbf{d}_{S'} = -\mathbf{s}_3$, it should be possible to find the coordinates of the image source, $\mathbf{p}_{S'} = [x_{S'}, y_{S'}]^T$, as the solution of the following linear system:

$$\mathbf{H}_{S'} \mathbf{p}_{S'} = \mathbf{d}_{S'}. \tag{5.2}$$

In real cases, we must consider the presence of noise that affects the estimates of the magnitude associated to each microphone and for each direction

of incidence. Although the estimate of the direction of arrival from the image source, through the main techniques of beamforming, is asymptotically not distorted, the presence of noise added to the single estimate an error factor that must be considered.

For that reason, we estimate $\mathbf{p}_{S'}$, as the (weighted) least squares (LS) solution of 5.2, given by:

$$\hat{\mathbf{p}}_{S'} = (\mathbf{H}_{S'}^T \mathbf{W}_{S'} \mathbf{H}_{S'})^{-1} \mathbf{H}_{S'}^T \mathbf{W}_{S'} \mathbf{d}_{S'}, \tag{5.3}$$

where $\mathbf{W}_{S'}$ is the weighting matrix assumed, for now, to be the identity matrix. This gives us an estimate of the image source position as shown in Figure 5.3 (a). In figure 5.1 we show a diagram of the alghorithm for the localization of single source.

2 Single reflector

Now we want to use the single image source estimation from the section before in order to obtain an estimation of the reflector endpoints that exploiting several acoustic images acquired by moving the direct source.

As observed earlier, not all microphones m_i sense the image source S' and the visibility of S' is limited by the dimensions of the reflector. In particular, the edges of reflector's visibility region lie on the planes that represent the endpoints A and B, as depicted in Figure 4.4 (b). So, our goal is now to estimate the edge rays. Knowing the image source position $\hat{\mathbf{p}}_{S'}$ and the last microphone(s) that sense it we can estimate the ray(s) on the edge of the visibility region, $\hat{\mathbf{a}}$ and/or $\hat{\mathbf{b}}$.

For the estimation of the last microphone(s) that sense the image source from the single reflector, we use the same procedure defined before for the $N_{S'}$ microphones that see S'.

Let us suppose that m_i is the obtained last microphone, and m_{i+1} the first one that are not illuminated by the image source (see Figure 5.3 (a)). Due to the sampling implicit on the use of a microphone array, we cannot say the exact point where the edge ray $\mathbf{a}$ intersect the line (the same applies for $\mathbf{b}$). However we expect that in average it intersect the array in mean point m_t, $\mathbf{p}_t = (\mathbf{p}_i + \mathbf{p}_{i+1})/2$. The estimated edge ray $\hat{\mathbf{a}}$ is then given by the intersection of planes $\overline{m_t}$ and $\overline{S'}$ as shown in Figure 5.3 (a):

$$\begin{aligned} \mathbf{x}_{S'}^T \mathbf{a} &= 0, \\ \mathbf{x}_t^T \mathbf{a} &= 0. \end{aligned} \tag{5.4}$$

From (5.4) $\mathbf{a} = [a_1, a_2, a_3]^T$ is a vector spanning the null space of $[\mathbf{x}_{S'}, \mathbf{x}_t]^T$. Supposing that $a_3 \neq 0$ (this can always be guaranteed changing the reference system) it can be normalized to 1. Indicating $\mathbf{G} = [\mathbf{p}_{S'}, \mathbf{p}_t]^T$ and $\mathbf{1} = [1, 1]^T$, we can find the remaining two coordinates, indicated with $\mathbf{a}^* = [a_1, a_2]^T$, as:

$$\begin{aligned} \mathbf{G}\mathbf{a}^* &= -\mathbf{1}, \\ \hat{\mathbf{a}}^* &= -\mathbf{G}^{-1}\mathbf{1}. \end{aligned} \tag{5.5}$$

Moving the source S we get a number of different acoustic images and as a consequence a number of edge rays $\hat{\mathbf{a}}^i$, $i = 1, ..., N_A$ and $\hat{\mathbf{b}}^j$, $j = 1, ..., N_B$ (see Figure 5.3 (b)). They lie on the edges of the visibility region, i.e. on the planes $\overline{A}$ and $\overline{B}$. In a similar way as done for the image source position in (5.3), we estimate the reflector endpoint positions using the LS method:

$$\begin{aligned} \hat{\mathbf{p}}_A &= (\mathbf{H}_A^T \mathbf{W}_A \mathbf{H}_A)^{-1} \mathbf{H}_A^T \mathbf{W}_A \mathbf{d}_A \\ \hat{\mathbf{p}}_B &= (\mathbf{H}_B^T \mathbf{W}_B \mathbf{H}_B)^{-1} \mathbf{H}_B^T \mathbf{W}_B \mathbf{d}_B. \end{aligned} \tag{5.6}$$

where $\mathbf{H}_A = [\mathbf{a}_1, \mathbf{a}_2]$, $\mathbf{H}_B = [\mathbf{b}_1, \mathbf{b}_2]$, $\mathbf{d}_A = -\mathbf{1}$, $\mathbf{d}_B = -\mathbf{1}$ and, for now, $\mathbf{W}_A = \mathbf{W}_B = \mathbf{I}$.

Notice however that in order to estimate A and B the edge rays $\hat{\mathbf{a}}^i$ and $\hat{\mathbf{b}}^j$ have to be visible by the microphone array. This requires a long or moving array and a moving source that illuminates the environment.

In Figure 5.2 we show a diagram that explains the algorithm for the estimation of a reflector endpoint.

3 Reflector line estimation

Now we describe two methods for the computation of the line $\mathbf{l}_R = (l_{R1}, l_{R2}, l_{R3})$ containing the reflector i.e the point in the RaySpace that represent the corresponding line. Note that this line is the result of classical estimation methods for the localization of reflective surfaces existing in the literature that do not allow to calculate the endpoints of the reflectors.

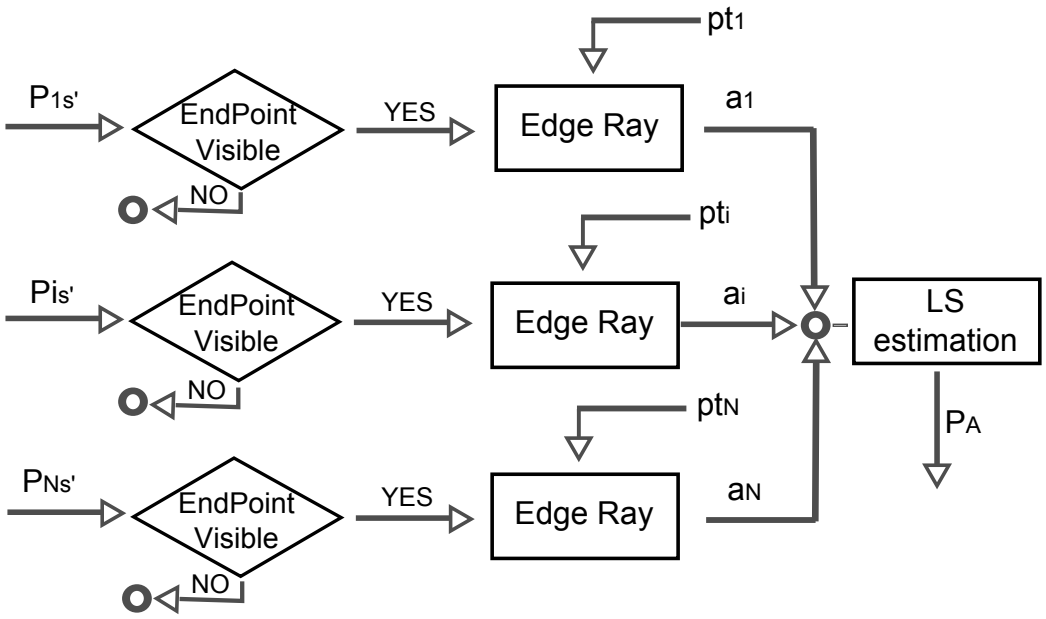


Figure 5.2: Diagram for the algorithm of estimation for an endpoint of a single reflector.

Once the endpoints A and B are estimated the line containing the reflector can be computed as the intersection of planes A and B in the RaySpace (i.e. joining the points A and B in the geometric space) as shown in Figure 5.3 (b):

$$\begin{aligned} \mathbf{x}_A^T \mathbf{l}_R &= 0, \\ \mathbf{x}_B^T \mathbf{l}_R &= 0. \end{aligned} \tag{5.7}$$

where $\mathbf{x}_A$ and $\mathbf{x}_A$ are the homogeneous coordinates for the endpoints of the reflector. However, as shown before, in order to obtain the estimates for A and B, we need a number of acoustic images for different position of direct source and a long or moving array in order to sense the edge rays of the reflector in different configurations.

The second method that we are going to present assumes to know the actual homogenous coordinates $\mathbf{x}_S$ of the direct source or an estimation of them obtained with the techniques previously seen. Let $\hat{\mathbf{x}}_{S'}$ be the estimate homogeneous coordinates of the image source for the first-order reflection from the given reflector. The reflector straight line can be obtained as:

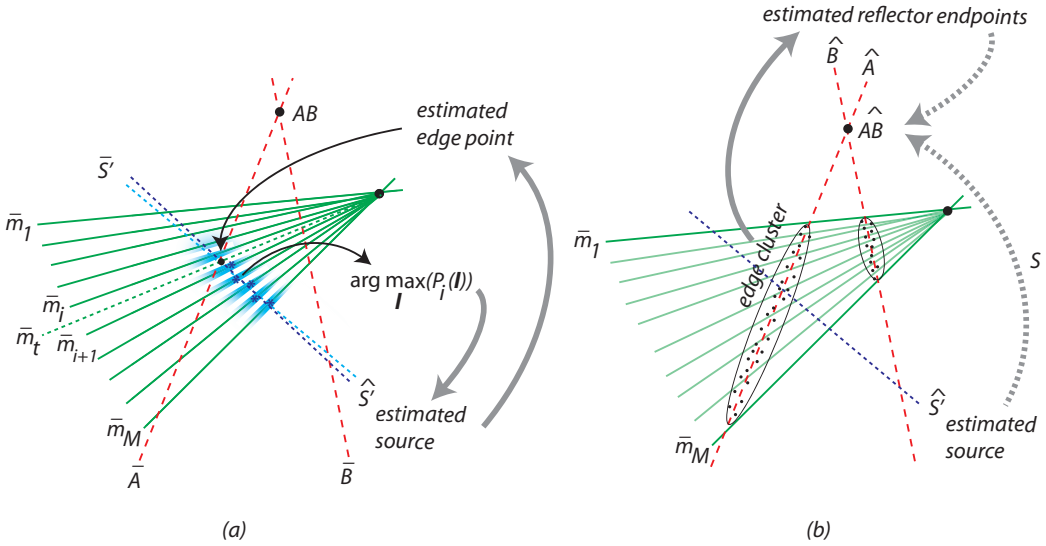


Figure 5.3: Overview of the estimation process: linear regression is used to estimate the source position (a) and reflector endpoints (b).

$$\mathbf{l}_R = \mathbf{x}_S - \hat{\mathbf{x}}_{S'} + \frac{1}{2}\left(\hat{\mathbf{x}}_{S'}^T G_1 \hat{\mathbf{x}}_{S'} - \mathbf{x}_S^T G_1 \mathbf{x}_S\right) G_2, \tag{5.8}$$

where:

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.9}$$

Note that, unlike the previous method, in order to obtain the estimation, we don't need a number of acoustic images for different position of direct source and a long or moving array in order to sense the edge rays because we need only the direct and image source of a single acquisition (or their estimation). However, it requires that the direct source is known or visible from the microphone array.

4 Multiple reflectors

In this section, we want to extend the previous methods to the cases in which multiple sources (direct source and image sources) are visible to the microphone array. That will allow us to estimate the position of multiple reflectors in a complex configuration.

However, a number of modifications is necessary in order to account for the presence of multiple sources and reflectors in the environment.

4.1 Localization of multiple sources

Suppose to have a configuration in which there are more than one source (direct sources and/or image sources) visible by the microphone array and their influence is not negligible. Our goal is to estimate their positions from a single acoustic image. Note that this problem occurs even when there is a single reflector but the signal emitted by the direct source appears to be significant and must be considered.

In presence of multiple sources we can not calculate simply the global maximum of the scanned image for each microphone m_i (i.e. the global maximum of $P_i(\mathbf{l})$) because a single microphone could acquire more than one signal from different sources. So we have to look for a number of local maxima of $P_i(\mathbf{l})$ $\forall i$ that have the acquired magnitude over a threshold, TS, defined once again as a percentage of the global maximum of the scanned image. We define R_S the set of all the rays that have these properties:

$$R_S = \left\{ \bar{\mathbf{l}} \in \mathbb{P}^2 | \ \exists i, \ \bar{\mathbf{l}} \in \arg loc \max_{\mathbf{l}}(|P_i(\mathbf{l})|), \ P_i(\bar{\mathbf{l}}) > TS \right\} \tag{5.10}$$

Now R_S contains all the rays acquired from all sources visible from the microphone array. In order to obtain the estimated locations using 5.3, we have to cluster the rays of R_S into sets representing contributions from different sources. This problem can be reformulated as follows: find the set of planes in the RaySpace (i.e. sources in the geometric space) that minimizes the sum of distances between the rays in R_S and the nearest plane (the total least square error). Exploiting the properties of the projective spaces, the problem will be handled in a suitable 2D reduced RaySpace (the acoustic image) . As a consequence the problem of finding planes reduces to a problem of finding lines that minimize the total least square error. A suitable reduced RaySpace can be obtained excluding the rays that are parallel to the microphone array. Now we present two different methods for finding optimal lines.

Method 1: comparison of clusters

First, we assume that there are only two sources visible from the array. Our aim is to find an exact algorithm for the definition of the corresponding straight lines. A brute force approach to the problem would be to define all possible partitions of the points available in two distinct sets, for each partition calculate the regression lines and take as solution the configuration with lower total least square error. This approach, although exact, has a computational complexity which makes it unenforceable except in academic cases. In fact, the number of partitions corresponds to cardinality of the power set for R_S and is therefore amounting to 2^n, where n represents the number of points considered. So we must search an approach that allows to decrease the number of configurations to compare.

The figure 5.4 a) shows a set of points and their two optimal straight lines S_1^* and S_2^*. In figure there are also two additional dotted lines which cross exactly in the middle of the angles between S_1^* and S_2^*. They are called decision bounders and they divide the space into four region that are assigned alternatively to one or the other line. So, if we define the two bounders, we can define the clusters of points and, from them, the optimal lines corresponding to the two sources.

In figure 5.4 b) we show also that we can traslate and rotate the given bounders in such a way as to obtain two new bounders that divide the points into the same clusters and such that:

1. The two bounders remains orthogonal;
2. One of the decision bounder goes through two input points;
3. The other decision bounder goes through an input point.

Now we can find all the possible bounders that have these properties. For all that bounders we can define the two clusters of points, compute the corresponding LS lines and calculate the total least square error. After that, we take the optimal lines (i.e the optimal sources locations) from the configuration with the lowest total least square error, [29].

The proposed algorithm has the advantage of ensuring the optimality of the solution obtained from it as it tests all the possible configurations. That algorithm has a complexity of $O(n^3)$ which is considerably better than the

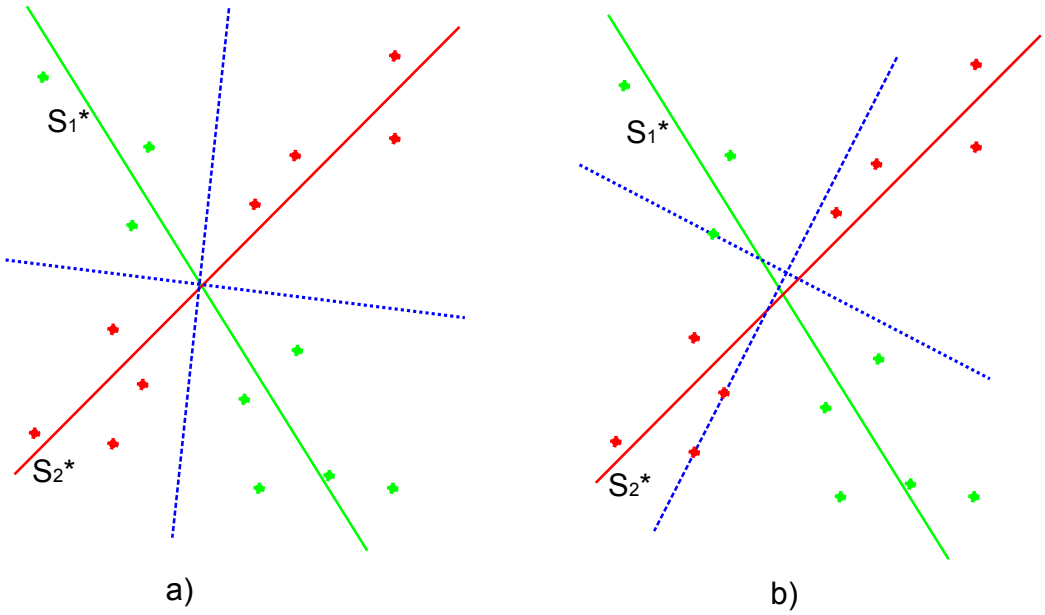


Figure 5.4: A set of rays, the two optimal lines and their bounders a); The corresponding translated and rotated bounders b).

brute force case. It could, however, be eccessive in cases where the number of points is high. Moreover, even if the algorithm can be extended to cases with more sources (for references see [29]), it will have a complexity ($O(n^{k^2})$ where k is the number of lines) that makes it unusable in not trivial cases. Another negative aspect of the algorithm is that it requires to know a priori the number of lines to search. We want an algorithm that automatically defines the number of sources seen from the microphone array.

These reasons justify the definition of an algorithm which, although not certifying the optimality of the solution obtained in the sense of the total least square error as the previous method, is very robust, has a computational cost suitable for more complex cases and defines automatically the number of sources seen. This method uses the Hough transform.

Method 2: robust Hough

The Hough transform is a technique that allows to recognize special configurations of points in an image representable as a matrix exploiting their parametric representation. In the case of interest, it is used for the identification of straight lines. These lines are parameterized according to length ρ

and angle θ of the vector perpendicular to them and having the origin in the reference center. So a line is defined by a point of coordinates (ρ, θ) in the parameters space. (Figure 5.5).

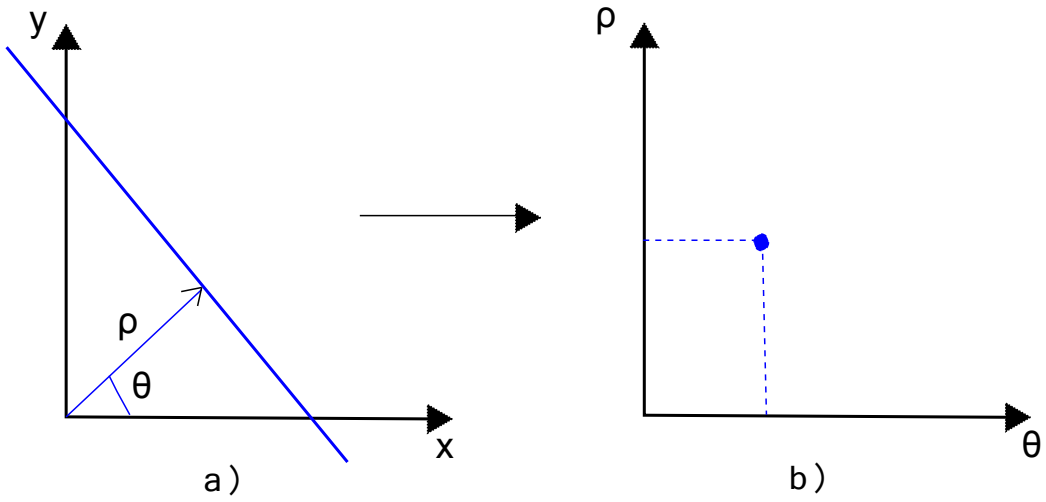


Figure 5.5: A line in the geometrical space a) and its representation in the parameters space b).

Since the Hough transform operates on a digital image represented as matrices of discrete samples, at first we perform a quantization of the continuous values of variables in R_S for the reduced RaySpace.

Then, for an arbitrary point on the image with coordinates (x_0, y_0), we define the lines that go through it as:

$$r(\theta) = x_0 \cos(\theta) + y_0 \sin(\theta), \tag{5.11}$$

where r (the distance between the line and the origin) is function of θ. This corresponds to a sinusoidal curve in the (r, θ) plane, which is unique for the point. If the curves corresponding to two points are superimposed, the location (in the Hough space) where they cross corresponds to the line (in the original image space) that passes through both points. More generally, a set of points that form a straight line will produce sinusoids which cross at the parameters for that line. Thus, the problem of detecting collinear points can be converted to the problem of finding concurrent curves.

In the case dealt, the points will not necessarily be perfectly aligned because they are affected by noise. Then we apply to the figure obtained by

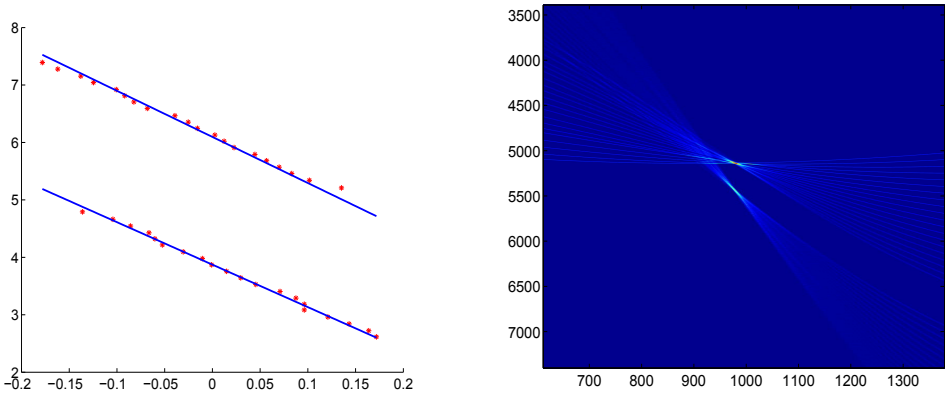


Figure 5.6: A 2D image of points with the lines defined by hough transform and the corresponding Hough space.

Hough transform in the parameter space a medium filter with appropriate width of the mask. If we know the number of sources to look for, we can pick up the corresponding number of peaks in the transform. Otherwise, if the number of visible sources is not defined a priori, it is possible to apply a threshold and select the number of local maxima above this threshold.

The straight lines obtained with the Hough transform do not necessarily minimize the total least square of points corresponding to the rays, as the transform works simply by points counter. For overcome this drawback, we put in series to Hough transform a k-means clustering algorithm, initialized with the solution of the previous step, using as distance the notion of classical Euclidean distance. This algorithm can be described as follows:

1. Join points to the nearest line and build the corresponding clusters;
2. For each cluster, calculate the least square line associated;
3. Check that the points are still assigned to the nearest line, if yes exit, if not repeat from 1.

The k-means clustering algorithm does not guarantee generally, when initialized randomly, global convergence. That explains the need to use as initialization a guess close to the solution, obtained through the peaks of the Hough transform for the image filtered in a suitable manner. In the full

version, the method has a low computational cost dominated by the research of the peaks relating to the Hough transform and complexity dependent by the levels of quantization of the parameters. In most cases it is also sufficient a single cycle of k-means clustering, as clusters are correctly defined by the output of the Hough transform. The algorithm results very robust and since the k-means algorithm is extremely fast, you can apply it multiple times with different initializations to confirm the goodness of the solution obtained.

4.2 Estimation of visibility region edges

As in the case of a single reflector (with a negligible direct signal), the visibility of image sources is limited by the dimensions of the corresponding reflector. In that case, the estimate of the first and last microphone acquiring the existing signal deriving from the only image source was simply done by calculating the total energy gained by each microphone, and finding if this energy exceeded a certain threshold defined as a percentage of the maximum intensity.

In the case with more reflectors, or in each case of more sources visible to the array, this technique is not applicable because the energy acquired by a microphone is the sum of the individual components that derived from different sources, which must be distinguished. We therefore propose two methods for the extension of previously presented algorithm.

Method 1: beamforming methods

The first proposed solution is to use the knowledge of the position of interested image source, estimated in the previous step, to calculate the energy incident from this direction through a beamforming method. Similarly as done to obtain the acoustic image, we have to window the array with a function w_i in order to test if the microphone m_i receives energy from the direction of the given image source.

Note that, given the position of the image source S' and the speed of sound c, it is possible to calculate the exact time delay for each microphone in the same window without using the farfield hypothesis (nearfield case). For example if we consider the two microphones m_i and m_j the delay in time τ_{ij} can be obtained as:

$$\tau_{ij} = \left(|S' - m_i| - |S' - m_j|\right)/c, \tag{5.12}$$

where $|\cdot|$ is the euclidean norm.
Such delays can be used to calculate the incident energy, through the choosen beamforming method, by the source of interest and allow to go beyond the classical assumptions of farfield of these methods.

It is usefull to underline that nevertheless once again the window w_i has to have limited aperture in order to reduce influence of neighbouring microphones. In any case, the method needs a window that considers more than one microphone because a single microphone senses only the overall signal and can not distinguish the various contributions. Consequently, the definition of the first and last acquiring microphones will be blurred of the amplitude of the used window.

Method 2: correlation matrix

We now propose an alternative solution for the same problem. This method derives from the observation that, since the knowledge of the estimated source allows to work with time delays in nearfield hypothesis, it is no longer necessary to use only microphones close to estimate the energy by beamforming. We define the directional energy matrix $E \in \mathbb{R}^{M\times M}$, where M is the number of microphones, such that for all couple of indices, E_{ij} is the the energy associated through a beamforming method from the given source to the two microphones by using time delays calculated as seen before. That is a symmetric matrix.

This matrix has the diagonal terms that are not-directional and can be very high even in microphones that do not acquire the signal from the studied source, due to the presence of other sources. To overcome this drawback we use instead the directional correlation matrix $C \in \mathbb{R}^{M\times M}$ defines as follow:

$$C_{ij} = \frac{E_{ij}}{\sqrt{E_{ii}}\sqrt{E_{jj}}}. \tag{5.13}$$

At this point, the method is based on considering the matrix C as a $2D$ image. The idea is to estimate the position of the last microphones that acquire the signal coming from the image source of the studied reflector, searching for the edges of the $2D$ image representing the correlation matrix

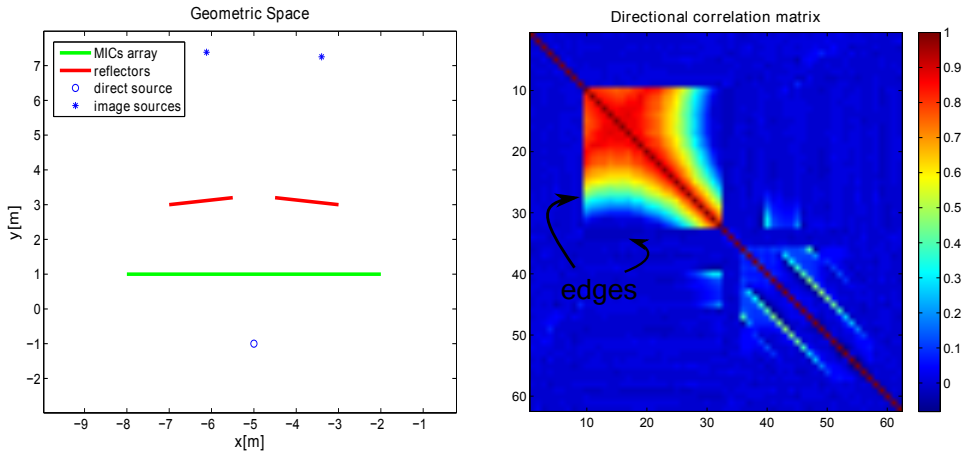


Figure 5.7: An example configuration and one of the corresponding directional correlation matrix.

of the acquired signal. In fact it is easy to see how the terms C_{ij} will be low if at least one of the two indices correspond to a microphone that is not affected by the signal of interest. In figure 5.7, we show an example of the use of this method for a simulated configuration. We can clearly see the edges corresponding to the last microphones that acquire the signal from the given image source.

The main disadvantage of this method is that it requires the calculation of the whole correlation matrix for each studied source and, if the number of microphones and the number of reflectors are high, it can have an high computational cost. The advantage is that it uses all the information available and the definition of the first and last acquiring microphones will not be blurred of the amplitude of the a window function.

4.3 Localization of multiple reflectors Endpoints

As in the case of a single reflector, we move the direct source in different positions in order to obtain different acoustic images of the examinated environment. For each acoustic image, we obtain a collection of the rays corresponding to visible endpoints for the reflectors in the studied configuration. These rays are grouped in different clusters, one for each visible endpoint, using the same techniques described for the source estimation.

After clustering the edge rays are used to estimate the edges of the given

visibility region in the ray space, i.e. estimate endpoints of the corresponding reflector. The procedure is the same as the one described for a single reflector case. In figure 5.8, we summarize the method.

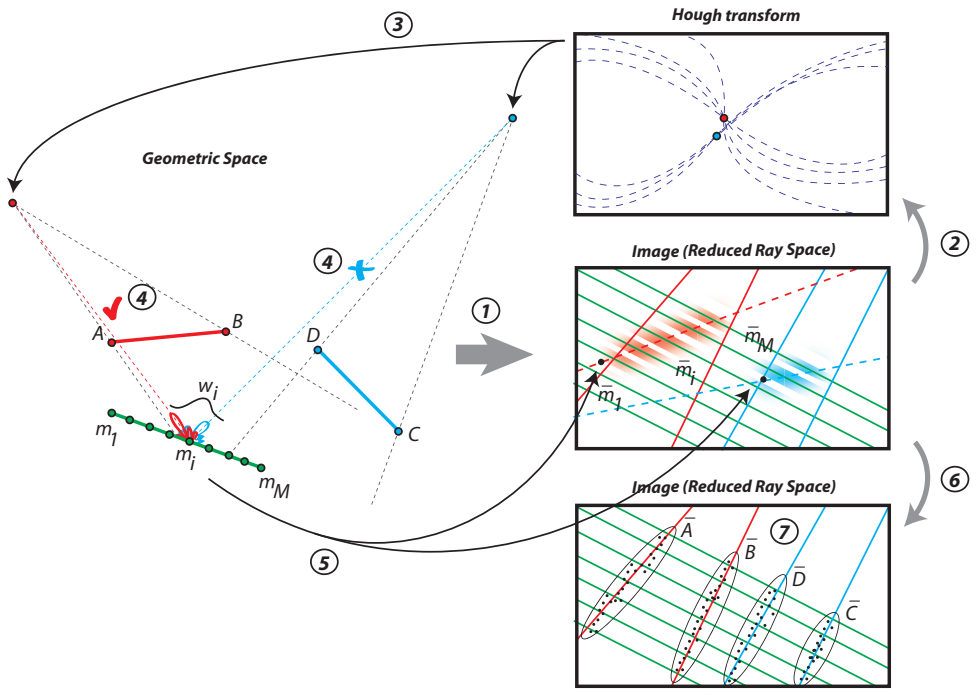


Figure 5.8: Estimation steps: 1) acoustic image is obtained from acoustic measurements; 2) Hough transform is used in order to individuate lines in the image; 3) positions of the image sources are estimated; 4) for each position m_i suitable techniques are used in order to test the visibility of the image sources; 5) if observable the edge rays are individuated and mapped to the RaySpace; 6) a number of edge rays are obtained moving the acoustic source; 7) clusters of edge rays are used to estimate reflector endpoints.

Chapter 6

Performance analysis

In this chapter we present an analysis of the properties of the estimation methods introduced before. In particular, in the first section we will derive expressions for second order statistics of estimates. This performance analysis is useful not only for system design but it will also be used to improve the estimation accuracy. The second section will be dedicated to the resolution analysis for the image source estimations.

1 Statistical analysis

1.1 Source localization

We start considering the accuracy of (5.1). The points $\hat{\mathbf{s}}^i$ correspond to maximum of $P_i(\mathbf{l})$ that is stricly correlated to the maximum of $P_i(\theta)$. The maximum of $P_i(\mathbf{l})$ are the estimated direction of arrival that we model as:

$$\hat{\theta}_i = \theta_i + \varepsilon_i, \tag{6.1}$$

where ε_i is the estimation error.

Discussion about beamforming methods is outside the scope of this work. Any method that takes into account the correlation between the sources (signals from various paths are attenuated and delayed versions of the same direct source signal) can be applied for this purpose. During this analysis we will suppose to know the performance of the given beamforming method and

as a consequence to know the $N_{S'} \times N_{S'}$ covariance matrix of DOA estimates $\mathbf{C}_\theta$. This covariance matrix will depend on the method used, windowing function w (i.e. the number of microphones used for the estimation), the spacing between microphones and signal frequency, directions of arrival and signal-to-noise ratio (SNR).

It is important to note that $\mathbf{C}_\theta$ is not diagonal as we estimate different θ_i using contributions from same microphones (shifting the window w) and thus θ_i will be correlated with θ_j, $j = i \pm 1, ..., N_w - 1$, where N_w is the window length. We suppose DOA estimates to be unbiased, i.e. ε_i to be zero mean.

Expanding $\hat{\mathbf{s}}^i = \mathbf{s}^i(\hat{\theta}_i)$ in first order Taylor series around the true value $\mathbf{s}^i(\theta_i)$:

$$\mathbf{s}^i(\hat{\theta}_i) \simeq \mathbf{s}^i(\theta_i) + \frac{\partial \mathbf{s}^i(\theta_i)}{\partial \theta_i}^T (\hat{\theta}_i - \theta_i),$$

we approximate the variances/covariances of estimated points as:

$$\begin{aligned} [\mathbf{C}_\mathbf{s}]_{ij} &= E\left[(\mathbf{s}^i(\hat{\theta}_i) - \mathbf{s}^i(\theta_i))(\mathbf{s}^j(\hat{\theta}_j) - \mathbf{s}^j(\theta_j))^T\right] \\ &\simeq \tfrac{\partial \mathbf{s}^i(\theta_i)}{\partial \theta_i}^T [\mathbf{C}_\theta]_{ij} \tfrac{\partial \mathbf{s}^j(\theta_j)}{\partial \theta_j}, \end{aligned} \tag{6.2}$$

where $E[\cdot]$ is the expectation operator, $[\mathbf{C}_\theta]_{ij}$ indicates the covariance between θ_i and θ_j, $\mathbf{C}_\mathbf{s}$ is the $3N_{S'} \times 3N_{S'}$ covariance matrix of estimated points and $[\mathbf{C}_\mathbf{s}]_{ij}$ indicates the 3×3 submatrix containing the covariances between points $\hat{\mathbf{s}}^i$ and $\hat{\mathbf{s}}^j$. From (4.11) with $k = 1$:

$$\frac{\partial \mathbf{s}^i(\theta_i)}{\partial \theta_i}^T = [\sin(\theta_i), \cos(\theta_i), -\sin(\theta_i)x_i - \cos(\theta_i)y_i]^T. \tag{6.3}$$

Writing (6.3) in (6.2) we have the desired covariance matrix.

Now we analyse the accuracy of the LS estimation of the image source position given by (5.3). In (5.2) both $\mathbf{H}_{S'}$ and $\mathbf{d}_{S'}$ are affected by errors, $\mathbf{H}^\varepsilon_{S'} = [\mathbf{s}^\varepsilon_1, \mathbf{s}^\varepsilon_2]$ and $\mathbf{d}^\varepsilon_{S'} = -\mathbf{s}^\varepsilon_3$. We model them as an additive noise:

$$\begin{aligned} \mathbf{H}_{S'}\mathbf{p}_{S'} &= \mathbf{d}_{S'} + \mathbf{n}_{S'}, \\ \mathbf{n}_{S'} &= -(\mathbf{s}^\varepsilon_1 x_{S'} + \mathbf{s}^\varepsilon_2 y_{S'} + \mathbf{s}^\varepsilon_3). \end{aligned}$$

Using (6.2), we calculate the covariance matrix of $\mathbf{n}_{S'}$ as:

$$\left[\mathbf{C}_{\mathbf{n}_{S'}}\right]_{ij} = [x_{S'}, y_{S'}, 1]\,[\mathbf{C}_{\mathbf{s}}]_{ij}\,[x_{S'}, y_{S'}, 1]^T. \tag{6.4}$$

The covariance of $\hat{\mathbf{p}}_{S'}$ is then given by:

$$\mathbf{C}_{\mathbf{p}_{S'}} = (\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{H}_{S'})^{-1}\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{C}_{\mathbf{n}_{S'}}\mathbf{W}_{S'}^T\mathbf{H}_{S'}(\mathbf{H}_{S'}^T\mathbf{W}_{S'}\mathbf{H}_{S'})^{-1}. \tag{6.5}$$

1.2 EndPoints localization

Now we consider the accuracy of the estimation for the endpoints. The estimate in (5.5) is affected by two independent types of errors.

1. Error $\epsilon_t = -(a_1 x_t^{\varepsilon} + a_2 y_t^{\varepsilon})$ caused by the sampling of the array (we assume that the detection method always detects correctly the microphones m_i and m_{i+1} for the given source). This error depends on geometry and is constant for given source (S') and array positions (m_t). As a consequence it adds a bias to the estimation of $\mathbf{a}^*$ that can be controlled decreasing the distance between microphones d;

2. Error $\epsilon_{S'} = -(a_1 x_{S'}^{\varepsilon} + a_2 y_{S'}^{\varepsilon})$ on image source position estimate caused by the blurring of the image. It is assumed to be zero mean with variance obtained from covariance matrix (6.5):

$$c_{\epsilon_{S'}} = E[\epsilon_{S'}^2] = \mathbf{a}^{*T}\mathbf{C}_{\mathbf{p}_{S'}}\mathbf{a}^*. \tag{6.6}$$

Then $\mathbf{G}\mathbf{a}^* = -\mathbf{1} + [\epsilon_{S'}, \epsilon_t]^T$ and the covariance matrix of $\mathbf{a}^{*i}$ is approximated as:

$$[\mathbf{C}_{\mathbf{a}^*}]_{ii} \simeq \mathbf{G}^{-1}\begin{bmatrix} c_{\epsilon_{S'}} & 0 \\ 0 & 0 \end{bmatrix}\mathbf{G}^{-T}. \tag{6.7}$$

Finally the covariances of estimates in (5.6) are obtained similarly to (6.5):

$$\begin{aligned} \mathbf{C}_{\mathbf{p}_A} &= (\mathbf{H}_A^T\mathbf{W}_A\mathbf{H}_A)^{-1}\mathbf{H}_A^T\mathbf{W}_A\mathbf{C}_{\mathbf{n}_A}\mathbf{W}_A^T\mathbf{H}_A(\mathbf{H}_A^T\mathbf{W}_A\mathbf{H}_A)^{-1}, \\ \mathbf{C}_{\mathbf{p}_B} &= (\mathbf{H}_B^T\mathbf{W}_B\mathbf{H}_B)^{-1}\mathbf{H}_B^T\mathbf{W}_B\mathbf{C}_{\mathbf{n}_B}\mathbf{W}_B^T\mathbf{H}_B(\mathbf{H}_B^T\mathbf{W}_B\mathbf{H}_B)^{-1}, \end{aligned} \tag{6.8}$$

where using (6.7):

$$\begin{aligned}[\mathbf{C}_{\mathbf{n}_A}]_{ij} &= \mathbf{p}_A^T\,[\mathbf{C}_{\mathbf{a}^*}]_{ij}\,\mathbf{p}_A,\\ [\mathbf{C}_{\mathbf{n}_B}]_{ij} &= \mathbf{p}_B^T\,[\mathbf{C}_{\mathbf{b}^*}]_{ij}\,\mathbf{p}_B.\end{aligned} \tag{6.9}$$

Unlike (6.4) the covariance matrices in (6.9) are diagonal as different edge rays are obtained with different observations, i.e. moving the source S, and are therefore uncorrelated. As already observed, due to the spatial sampling of the microphone array, the single edge rays $\hat{\mathbf{a}}^i$ and $\hat{\mathbf{b}}^j$ are biased. As a consequence the estimates of reflector endpoints $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_B$ are biased in general. However, the bias on edge rays can be modelled as uniformly distributed between the two microphone samples (last one that sense the image source and the first one that does not) and thus $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_B$ can be considered asymptomatically unbiased (for N_A and N_B going to infinity).

1.3 Reflector line estimation

We take into account the analysis of variance/covariance for the estimations of the line containing the reflector in the two modalities described in the chapter 5. At first, we consider the analysis of (5.7).

We suppose to know the covariance matrices for the endpoints, $\mathbf{C}_{\mathbf{p}_A}$ and $\mathbf{C}_{\mathbf{p}_B}$. We suppose also that for the reflector straight line $\mathbf{l}_R = (l_{R1}, l_{R2}, l_{R3})$, $l_{R3} = 1$. The assumption is legitimate in any case where this parameter is not zero. This can be obviated changing the reference system. Now we define $\mathbf{H}_R = [\mathbf{p}_A, \mathbf{p}_B]^T$, $\mathbf{l}_{R'} = (l_{R1}, l_{R2})$ and $\mathbf{d}_R = [-1, -1]^T$. So, the system (5.7) can be formulated as:

$$\mathbf{H}_R\mathbf{l}_{R'} = \mathbf{d}_R. \tag{6.10}$$

In (6.10), $\mathbf{H}_R$ is affected by errors, $\mathbf{H}_R^\varepsilon = [\mathbf{p}_A^\varepsilon, \mathbf{p}_B^\varepsilon]$ and. We model it as an additive noise:

$$\mathbf{H}_R\mathbf{l}_{R'} = \mathbf{d}_R + \mathbf{n}_R, \tag{6.11}$$

$$\mathbf{n}_R = \begin{bmatrix} x_A^\varepsilon l_{R1} + y_A^\varepsilon l_{R2} \\ x_B^\varepsilon l_{R1} + y_B^\varepsilon l_{R2} \end{bmatrix}$$

Using $\mathbf{C}_{\mathbf{p}_A}$ and $\mathbf{C}_{\mathbf{p}_B}$, we can calculate the covariance matrix of $\mathbf{n}_R$ under the assumption that the coordinates for one endpoint are indipendent from the coordinates of the other (i.e that there aren't image sources that sense both

the endpoints). If this assumption is not verified, we must also consider the cross-covariances.

Now, we analyse the variance/covariance for the reflector straight line calculated as in (5.8). We expand the estimation $\hat{\mathbf{l}}_R = \mathbf{l}_R(\hat{\mathbf{x}}_{S'})$ in first order Taylor series around the true value $\mathbf{l}_R(\mathbf{x}'_S)$ and we approximate the variances/covariances of estimated points as:

$$\mathbf{C}_{\mathbf{l}_R} \simeq \frac{\partial \mathbf{l}_R(\hat{\mathbf{x}}_{S'})}{\partial \mathbf{x}_{S'}}^T \mathbf{C}_{S'} \frac{\partial \mathbf{l}_R(\hat{\mathbf{x}}_{S'})}{\partial \mathbf{x}_{S'}}, \tag{6.12}$$

From (6.10) we have that:

$$\frac{\partial \mathbf{l}_R(\hat{\mathbf{x}}_{S'})}{\partial \mathbf{x}_{S'}} = \begin{bmatrix} -1 & 0 & x_{S'} \\ 0 & -1 & y_{S'} \end{bmatrix} \tag{6.13}$$

1.4 ML estimation

The previous performance analysis is useful for different reasons. At first, that analysis can be used to design the acquisition system in order to obtain desired variance/covariance properties. We will show an example in the next chapter. Furthermore, it can also be used to improve the estimation accuracy. In fact the weighted LS estimates (5.3) and (5.6) become maximum likelihood (ML) estimates for $\mathbf{W}_{S'} = \mathbf{C}_{\mathbf{n}_{S'}}^{-1}$, $\mathbf{W}_A = \mathbf{C}_{\mathbf{n}_A}^{-1}$ and $\mathbf{W}_B = \mathbf{C}_{\mathbf{n}_B}^{-1}$ and the idea is to estimate these weights from the previous results. In fact the 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 (6.4) and (6.9)) and therefore they cannot be known a priori (they represent the goal of our estimate). Nonetheless the LS estimates can approach ML estimates iteratively. The algorithm proceeds as follows:

1. obtain initial estimates $\hat{\mathbf{p}}_{S'}$, $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_B$ with $\mathbf{W}_{S'} = \mathbf{I}$, $\mathbf{W}_A = \mathbf{I}$ and $\mathbf{W}_B = \mathbf{I}$;
2. estimate $\hat{\mathbf{C}}_{\mathbf{n}_{S'}}$, $\hat{\mathbf{C}}_{\mathbf{n}_A}$ and $\hat{\mathbf{C}}_{\mathbf{n}_B}$ using the estimations from the previous point;
3. obtain new estimates $\hat{\mathbf{p}}_{S'}$, $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_B$ with $\mathbf{W}_{S'} = \hat{\mathbf{C}}_{\mathbf{n}_{S'}}^{-1}$, $\mathbf{W}_A = \hat{\mathbf{C}}_{\mathbf{n}_A}^{-1}$ and $\mathbf{W}_B = \hat{\mathbf{C}}_{\mathbf{n}_B}^{-1}$;
4. repeat from step 2.

The iterative procedures stops when the desired iteration number is reached. Simulations will show that the algorithm gets close to the ML performance in just one iteration with a considerable gain with respect to LS estimates.

2 Resolution of the method

We take into account the problem of resolution i.e the ability of the method to distinguish two reflectors. We make the assumption that the first order reflections are visible. So we reformulate the problem as follows: be able to separate the contributions of two image sources corresponding to two reflectors, i.e. distinguish the two peaks in the spatial pseudospectrum.

Given the DOA estimation technique we suppose to know the angular resolution of the method $\Delta\alpha_{\min}(\theta)$ (i.e the minimum angle to resolve two different peaks). This resolution will depend on the method used, windowing function w (i.e. the number of microphones used for the estimation),and the number and values of the signal frequencies that are used to apply the beamforming method. It will be also function of the directions of arrival (as shown in the definition) because we expect to have a better resolution for image sources in front of the array than decentralized with respect to it.

In order to analyze the resolution of the method, we suppose that we can define exactly the direction of arrival of the rays. At first, we suppose also to have a single microphone (a single windowed set of microphones, Figure 6.1) that sense the signals that arrive from the two different reflectors. Let define with $\mathbf{p}_{S1'}$ and $\mathbf{p}_{S2'}$ the euclidean coordinates for the image sources for the two reflectors and $\mathbf{p}_S$ the euclidean coordinates for the direct source. Let define also $\mathbf{l}_{R1}$ and $\mathbf{l}_{R2}$ the homogeneous coordinates for the two reflectors straight lines. We have that the following formulas can be used to write the euclidean coordinates for the image sources as functions of the coordinates of the direct source and the reflectors straight lines:

$$\mathbf{p}_{S1'} = G_2 * \left(I_3 - 2\frac{G_1\mathbf{l}_{R1}\mathbf{l}_{R1}^T}{\mathbf{l}_{R1}^T G_1 \mathbf{l}_{R1}} \right) \mathbf{p}_S \tag{6.14}$$

$$\mathbf{p}_{S2'} = G_2 * \left(I_3 - 2\frac{G_1\mathbf{l}_{R2}\mathbf{l}_{R2}^T}{\mathbf{l}_{R2}^T G_1 \mathbf{l}_{R2}} \right) \mathbf{p}_S \tag{6.15}$$

with

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{6.16}$$

Now the minimum requirement for separation is that the angle formed by $\mathbf{p}_{S1'}$, $\mathbf{p}_{S2'}$ and the central microphone of the window is greater than the resolution of the method of beamforming $\Delta\alpha_{\min}(\theta)$. Then, using the law of the cosine and denoting by $\mathbf{m}$ the euclidean coordinates of the microphone, we have the following condition:

$$\arccos\left(\frac{|\mathbf{p}_{S1'} - \mathbf{m}|^2 + |\mathbf{p}_{S2'} - \mathbf{m}|^2 - |\mathbf{p}_{S1'} - \mathbf{p}_{S2'}|^2}{2\,|\mathbf{p}_{S1'} - \mathbf{m}|\,|\mathbf{p}_{S2'} - \mathbf{m}|}\right) > \Delta\alpha_{min}(\theta), \tag{6.17}$$

where as θ we use the mean of the direction of arrival for the two image sources and $|\cdot|$ is the euclidean norm.

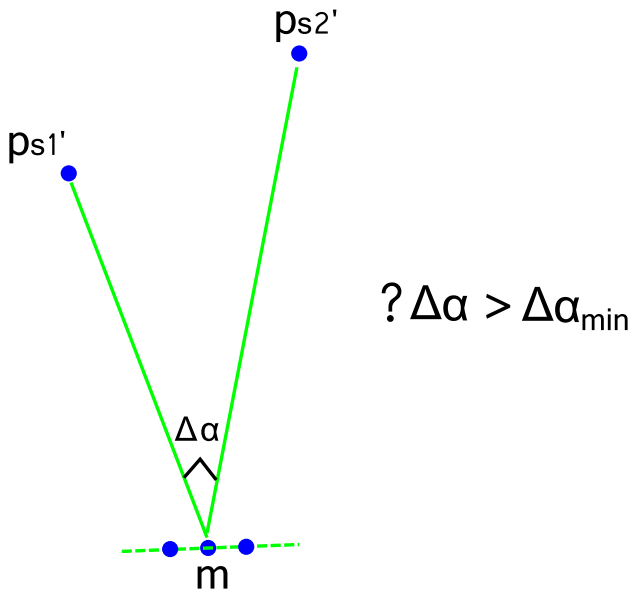


Figure 6.1: Resolution problem: the angular distance of the image sources have to be higher than the method resolution.

Exploiting the fact that the restriction of the cosine function between

$0 - \pi$ is monotonically decreasing, we can rewrite (6.17) as:

$$\frac{|\mathbf{p}_{S1'} - \mathbf{m}|^2 + |\mathbf{p}_{S2'} - \mathbf{m}|^2 - |\mathbf{p}_{S1'} - \mathbf{p}_{S2'}|^2}{2\,|\mathbf{p}_{S1'} - \mathbf{m}|\,|\mathbf{p}_{S2'} - \mathbf{m}|} < \cos\left(\Delta\alpha_{min}(\theta)\right). \tag{6.18}$$

Let us now consider the real case in which there will be a number of microphones M with euclidean coordinates $\mathbf{m}_i$ with $i = 1, ..., M$. In this case, the image sources corresponding to the reflectors are certainly resolved if there are microphones that sense only the first image source and microphones that sense only the second. Otherwise, the previous condition must be tested on each set of windowed microphones to verify the existence of microphones that can distinguish the two reflectors.

Chapter 7

Simulations and Experiments

In this chapter we will present simulations and experiments in order to validate the estimation techniques introduced in the previous chapters. The simulations have also the purposes to show a use of the statistical analysis for the design of a system for reflector estimation with a desired accuracy, to analyse the performance for different values of SNR and to show the gain obtained using the iterative algorithm instead of LS. Experiments were performed to demonstrate the effectiveness of the methods in cases of real acquisitions.

1 Simulations

1.1 Number of Microphones

In the first test we want to design a system for reflector estimation using the variance/covariance analysis described before. For simplicity we suppose to have an array with fixed microphone spacing $d = 10cm$. The design variable is the number of microphones M. At disposal we have two soundbars each one composed by 5 loudspeakers spaced by $20cm$. The soundbars, placed behind the array at distance $2m$, are used to excite the environment with a white noise (see Figure 7.1).

The project scope statement requires the system to be able to estimate the position of a $1.5m$ long reflector placed $2m$ in front of the array with $\text{var}(x_D) = \text{var}(x_C) \leq 4 \cdot 10^{-6}$ and $\text{var}(y_D) = \text{var}(y_C) \leq 4 \cdot 10^{-4}$ at $SNR = 20dB$.

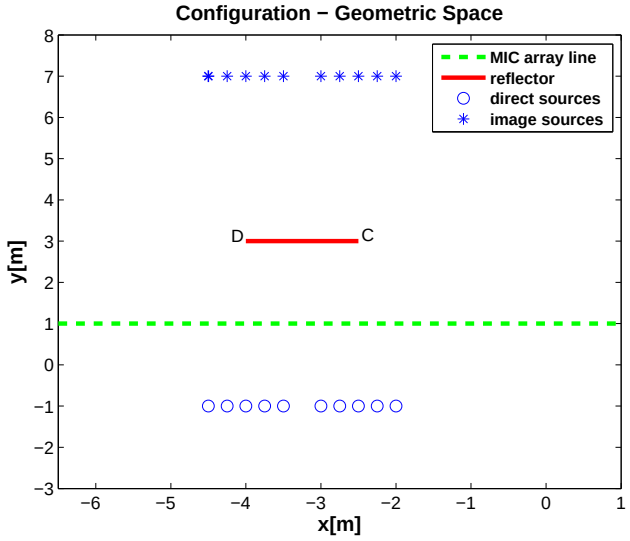


Figure 7.1: System configuration: microphone array (green line) is used to estimate the position of the test reflector (red line) excited by 10 acoustic sources (blue circles).

Due to the $10cm$ microphone spacing we use the 3-sample rectangular window w and the wideband Capon method [27] to perform beamforming. Now we use the analytical expressions for the covariance of the reflector endpoints in order to obtain the expected variances as a function of the interested variables. We use the covariance matrix $\mathbf{C}_\theta$ estimated from simulations to performe the analysis. We repeat the calculus for different values of the number of microphones (the design variable). It is always assumed that the array of microphones is centered relative to the reflector. Furthermore, the number of microphones varies between 14 and 36 because it can be deduced that 14 is the minimum number of microphones (under the specifications defined above) necessary for estimating the endpoints and beyond 36, additional microphones would acquire only noise (they would be outside the region affected by the image source).

In Figure 7.2, we show the results of theoretical estimation for the variance of the endpoints. We show the plots for the complete range of values for the number of microphones and the details on the interested regions. Note that it is possible to consider only one endpoint because the system is completely simmetric. The greatest benefit in terms of reducing the variance of the estimate by adding microphones occurs when the number of microphones

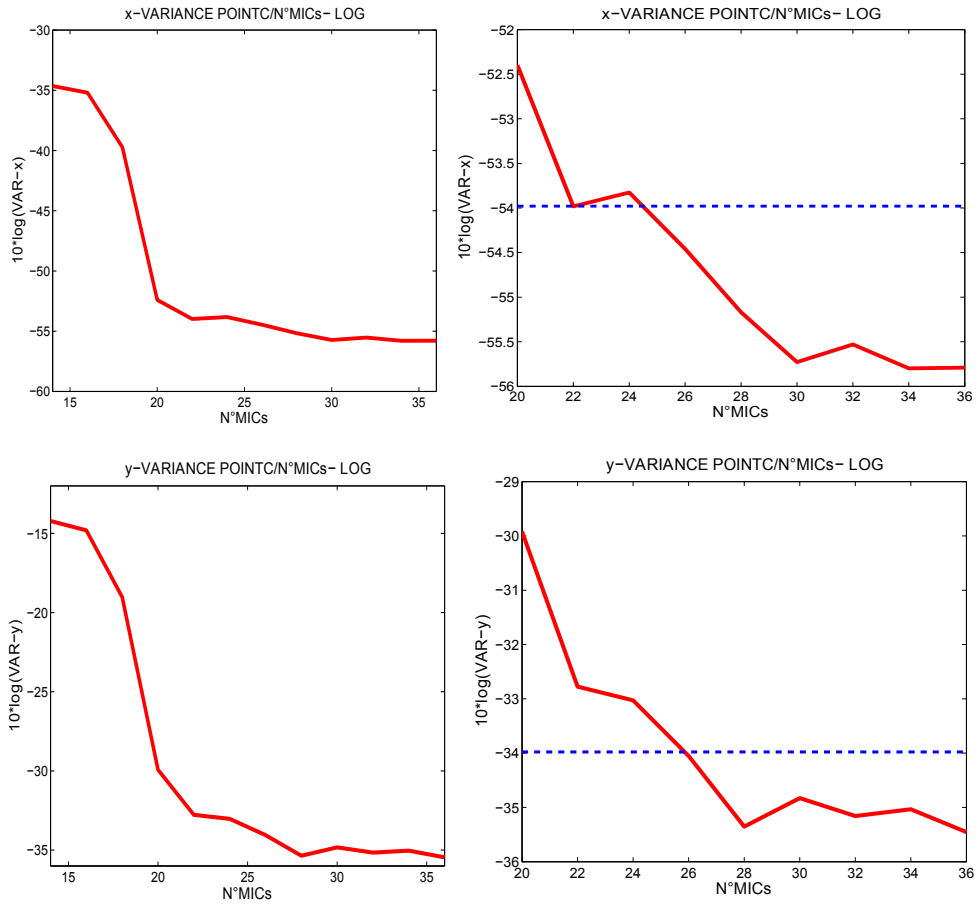


Figure 7.2: Endpoint estimation variance in function of the number of microphones (red) and the desired variance (blue); being the system completely symmetric only the data for point C is visualized.

is still low. Moreover there is not a regular behavior of the solution to the variation of the number because there are additional microphones that are used only for the estimation of the image source and microphones that increase the number of endpoints perceived.

Examining the figure the optimal (lower) number of microphones M that satisfies the specifications is found to be 26. Finally we perform 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 are reported in the following table where it can be observed how the simulated values confirm the theoretical analysis:

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

Table 7.1: Results for the system design; Theoretical and simulation variances.

1.2 Covariance/SNR

In the second test we analyse the performance for different values of SNR and the gain obtained using the iterative algorithm instead of LS. The array used in simulation has 25 microphones. We use a wideband capon method to perform the beamforming. In particolar, the product of an harmonic and geometric mean is performed on the beamforming values for all the frequencies analyzed. The geometric mean causes the suppression of peaks that are not present in all the frequencies and allows to use a maximal frequency that is more than the theoretical one (to prevent aliasing). This improves the resolution of the method (i.e reduces the width of the peaks). This method is used in all the following simulations and experiments. The configuration is shown in Figure 7.3 (a). Figure 7.3 (b) shows an example of the obtained acoustic images (visualized in the reduced ray space with $l_1 = 1$) for a single source position and $SNR = 20dB$. We can note as the direct source allows to illuminate the reflector that is seen by the microphone array. The acquired part of the RaySpace shows also that both the endpoints for the reflector are visible. The gap between the end of the acquired area and the microphone array endpoints derives from the width of windows and it corresponds to one microphone (a window of 3 microphones is used). The reflector estimation obtained moving the source across 21 different positions in space is shown in Figure 7.3 (a).

The Figure 7.4 shows the variances of endpoint estimates for different values of SNR. In particular: 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 algorithm; and the theoretical variance for the ML technique.

The figure puts in evidence a number of aspects:

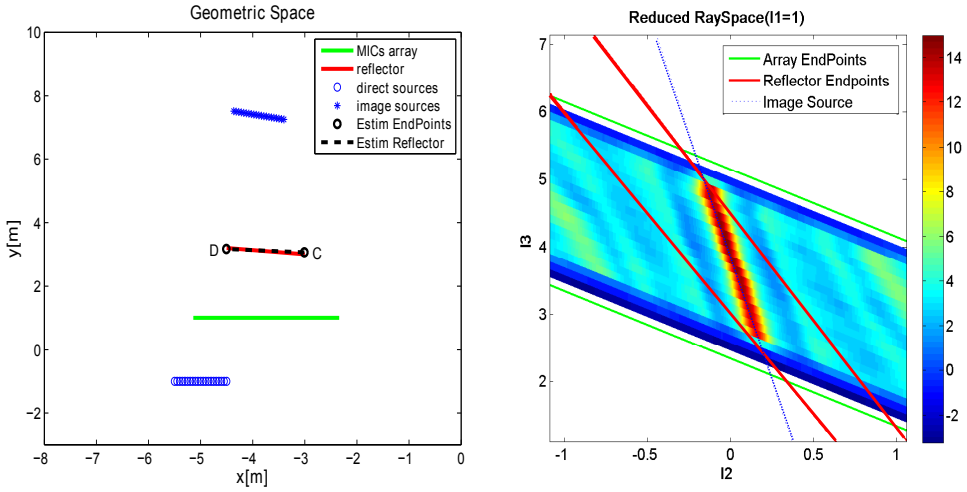


Figure 7.3: The configuration in the geometric space a) with estimation example (black dotted line); example of the measured acoustic image shown in the reduced RaySpace b).

- the theoretical variance approximates well the variance obtained using simulations. Moreover, the error estimation is higher for low values of the signal-to-noise ratio;
- the iterative algorithms achieves a significant gain with respect to LS technique (i.e. a reduction of the variances for the endpoints estimation);
- one cycle of the iterative algorithm gets close to the performance of the ML technique and as a consequence the gain obtained by successive iterations is negligible.

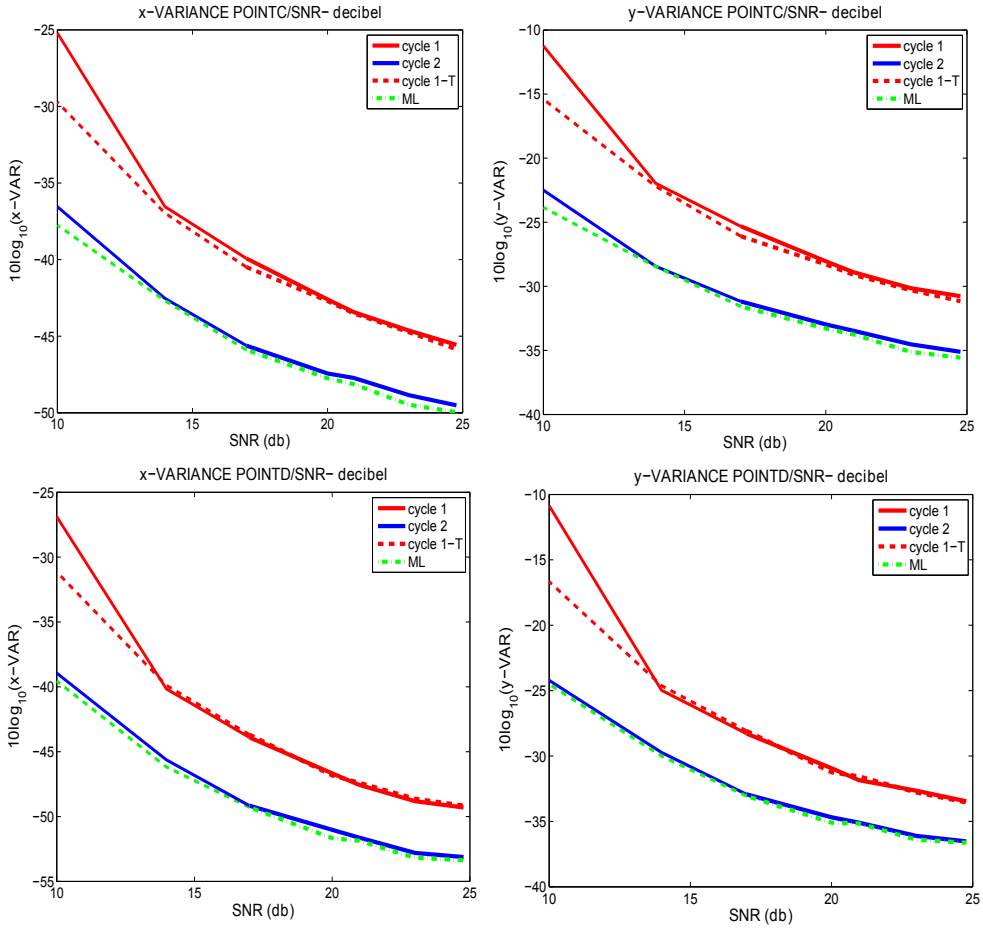


Figure 7.4: The theoretical variance obtained using analytical expressions for the LS technique (red dotted line); the simulation variance estimated from 1000 independent realizations for the LS technique (red line); the simulation variance for one cycle of the iterative algorithm (blue line); and the theoretical variance for the ML technique (green dotted line).

1.3 Multiple reflectors examples

Now we show some results obtained using configurations with multiple reflectors. The simulation examples for the cases with two and three reflectors are shown in Figures 7.5 and in Figure 7.6. In figure 7.5 a) and 7.6 a) we show the configurations in the geometric space and the estimations obtained through simulations. In the 7.5 b) and 7.6 b) we show the corresponding solution in the reduced RaySpace.

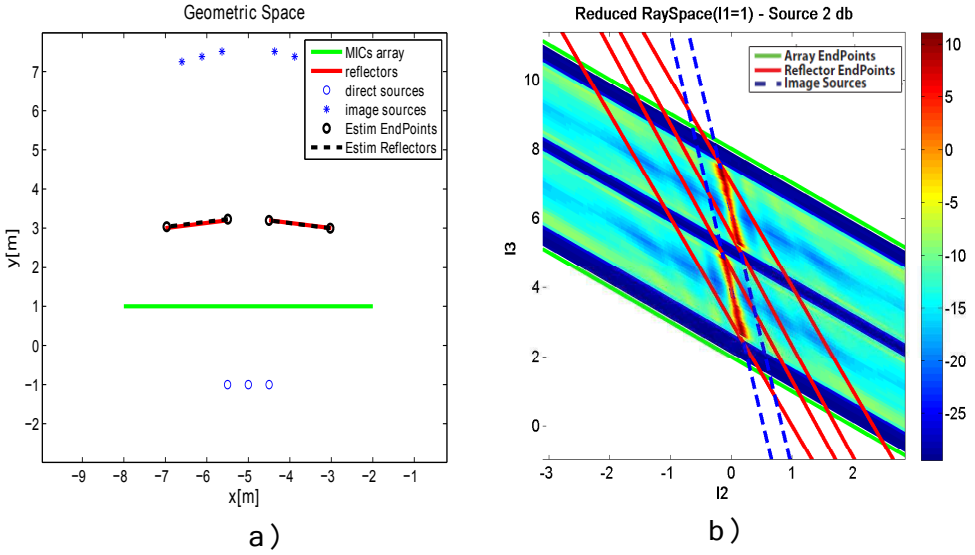


Figure 7.5: Example with two reflectors. The configuration in the geometric space a) with estimation examples (black dotted lines); example of the measured acoustic image shown in the reduced RaySpace b).

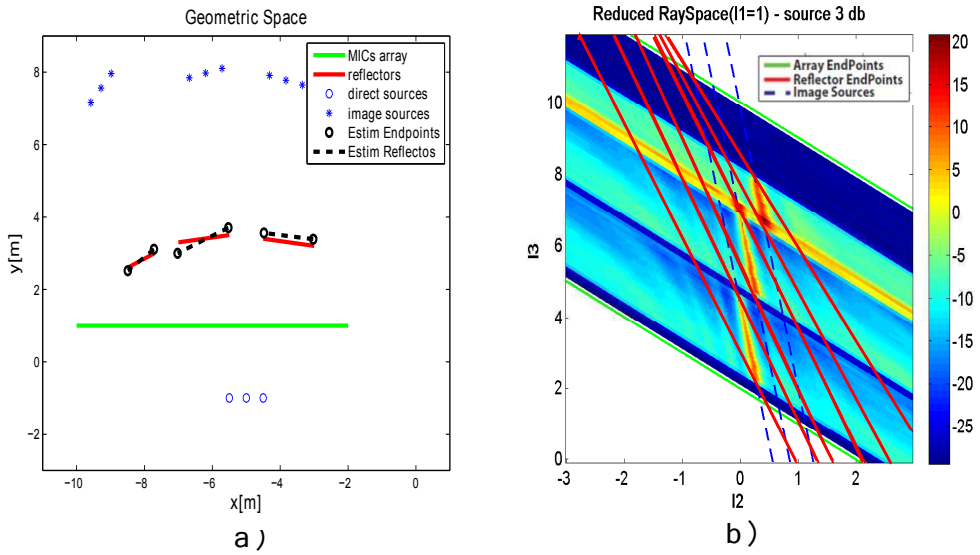


Figure 7.6: Example with three reflectors. The configuration in the ray space a) with estimation examples (black dotted lines); example of the measured acoustic image shown in the reduced RaySpace b).

In Figure 7.7, we show the the Hough transforms for images in Figures 7.5 b) and 7.6 b) for a fixed direct source. We can see the peaks of the Hough

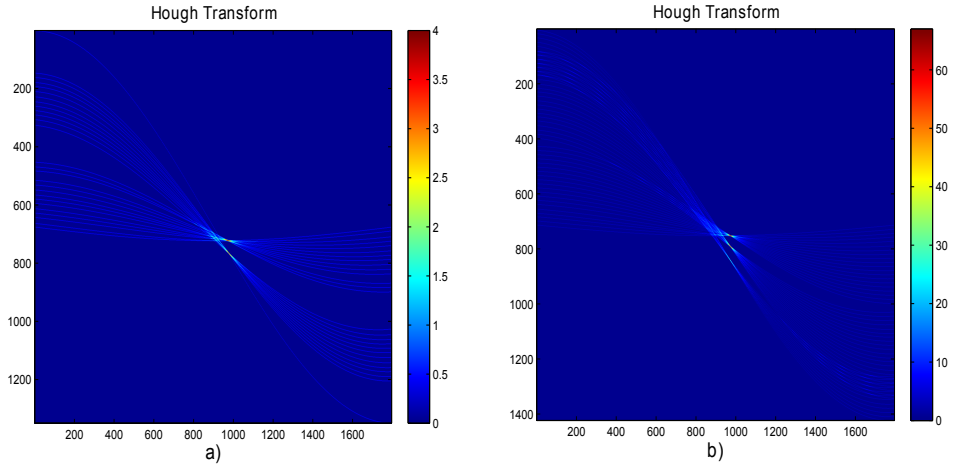


Figure 7.7: a) Hough transform of image in Figure 7.5 b) Hough transform of image in Figure 7.6.

transform that correspond to the image sources of the reflectors.

2 Experiments

2.1 Experimental setup

Finally we perform experiments in order to test the methods through real acquisitions. The acquisitions were performed in the Sound and Music Computing Lab at the Politecnico di Milano inside an anechoic chamber. An anechoic chamber is a room designed to stop reflections of sound. It is also insulated from exterior sources of noise. These two properties are very usefull to perform experiment and test the methods of estimation. The walls of the configurations are created using reflective panels. The direct source is reproduced using a directional loudspeaker (see Figure 2.1). The source emits in all the experiments a noise for an acquisition time of 2 seconds.

(a) The loudspeaker. (b) The microphone array.

Figure 7.8: Experimental equipment for real acquisitions.

The first experiments are permormed using a microphone array of 13 microphones (see Figure 2.1). The frequency of acquisition is in all the experiments $44100Hz$. It is important to note that the microphone array is linear but not uniform (the distance between microphones is variable). This does not appear to be a problem for the method under consideration because the developed methods can be easily adapted to this type of array. Obviously the use of such array means that we expect to have higher errors made in the approximation of the final point of acquisition if the ray interest a part of the array with higher distance between the microphones.

The last experiment is performed using a configuration that represent a closed environment. In this case we use a uniform array of 16 microphones. The frequency of acquisitions is still $44100Hz$.

2.2 Localization of acoustic source

In the first experiment we test the system for localization of acoustic sources. The loudspeaker emitting a white noise is placed in front of the array as shown in Figure 7.9 a). The acoustic image acquired by the array is shown in Figure 7.9 b). In Figure 7.9 the actual and estimated source positions are visualized in both geometric space a) and RaySpace b).

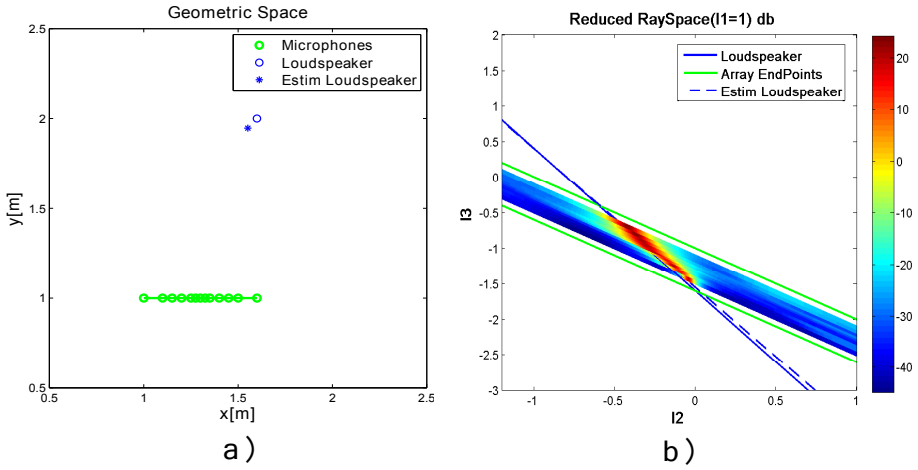


Figure 7.9: Configuration in the geometric space a) and the corresponding acoustic image b) with the estimated source position.

2.3 Reflector line estimation

Next we place a $60cm$ long reflector in front of the array. The experimental setup is shown in Figure 7.10. The presence of the reflector causes reflections to appear. The array is then used to estimate the position of both real and image loudspeakers. As we can note in Figure 7.11 (b), in the acquired RaySpace we can see both the planes (lines if shown in the recuced RaySpace) corrisponding to the direct source from the loudspeaker and the image source generated by the first reflection of the single reflector and they are both

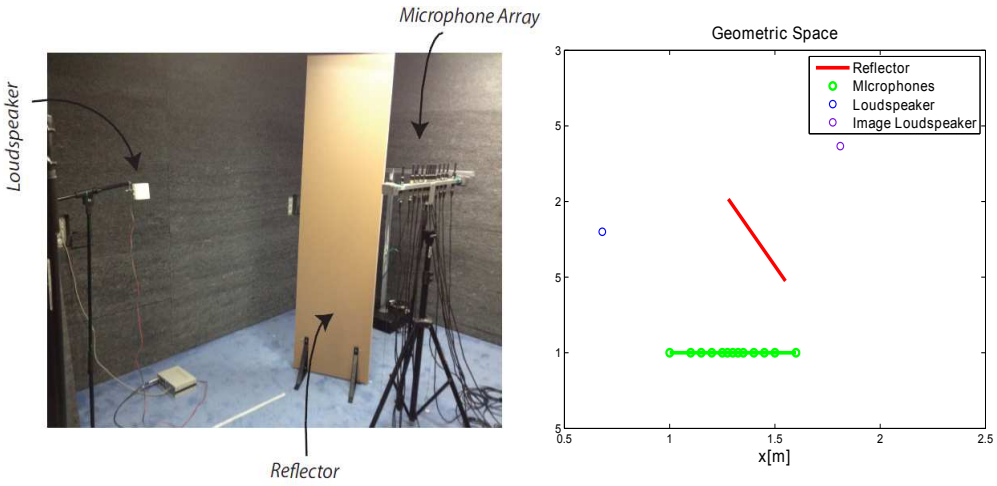


Figure 7.10: Experimental setup for the first configuration.

estimated. They can then be used to estimate the line on which the reflector lies. The results of the experiment are shown in Figure 7.11 a) and b) and are confronted with simulation results shown in Figure 7.11 c) and d).

We remember that the straight line on which the reflector lies is the only result of the previous methods for reflector estimation. This experiment shows also as the desribed method can be used to estimate not only the geometry of the environment but also the configuration of the acoustic system (the position of the direct source) if it is not a data for the problem. This will be possible whenever the direct source is visible by the microphone array.

2.4 EndPoint estimation

Finally the reflector endpoints can be estimated if a number of edge rays is observed (theoretically ≥ 2). This can be achieved using a long array and/or moving the acoustic source. We use for the estimation of reflector endpoints 10 different positions for the loudspeaker as shown in Figure 7.12 (a). The microhones array position is on the contrary fixed.

It is easy to see that only one endpoint is visible by the array from the image sources generated by the loudspeaker positions. The other endpoint can not be estimated. The Figure 7.12 b) shows the estimated image sources and the corresponding edge rays in the reduced RaySpace. It is also shown the line corresponding to the visible endpoint. The figure 7.12 a) shows the

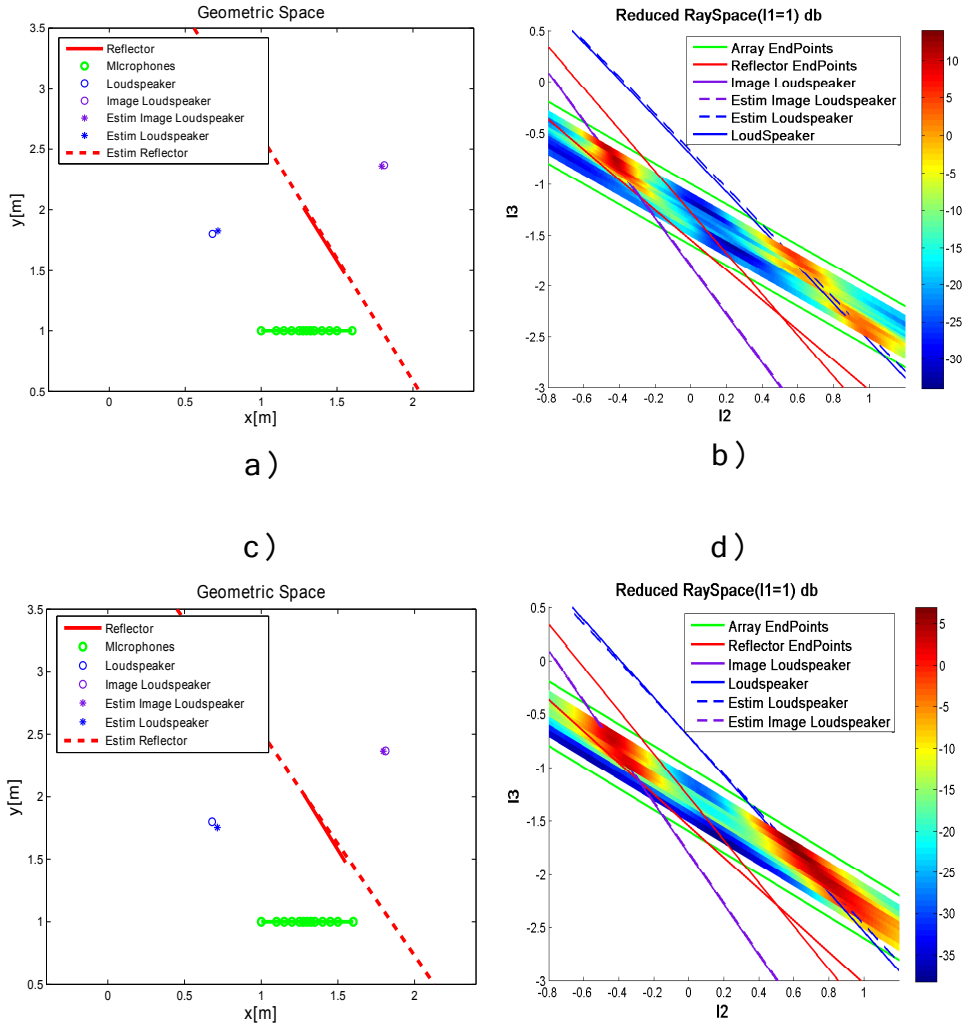


Figure 7.11: Configuration and experiment estimates in the geometric a) and ray space b); configuration and simulation estimates in the geometric c) and ray space d).

results in the geometric space.

2.5 Room estimation

Now we take into account one case of a more complex environment. The configuration is shown in Figure 7.14. This configuration is used to test the methods because it appears to be problematic for classical approaches in the literature. In particular, the estimate of the blue wall in the figure

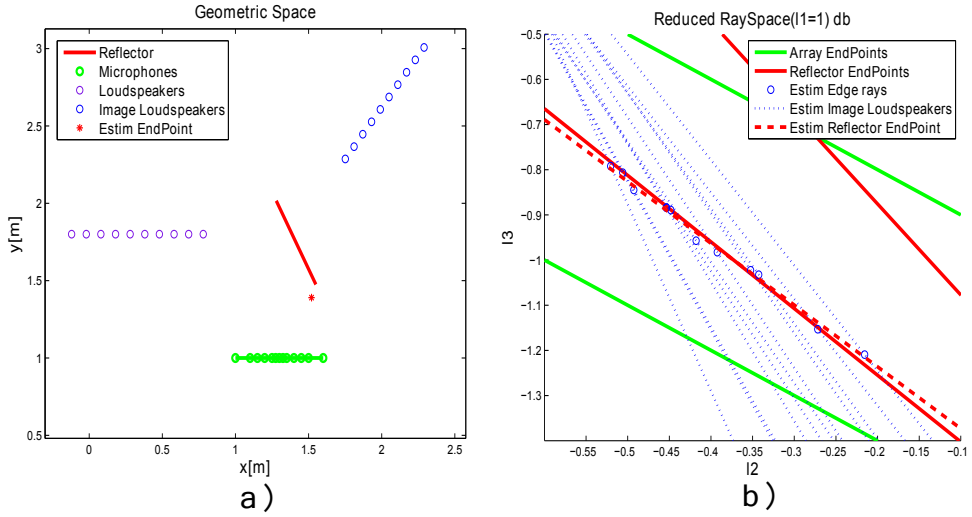


Figure 7.12: Source positions in the geometric space a) and corresponding edge ray estimates in the ray space b) used for endpoint estimation.

is difficulty executable by methods that do not take into account partial visibility and occlusions of the reflector or that use hardware systems with limited dimension.

The Figure 7.14 shows the visibility of the blue wall with respect to the array of microphones in function of the position of the direct source. In particular, the direct sources in the red area do not allow to see this wall, in the blue allow a partial view of the wall (and thus to see the corresponding end point) and in the green allow a total view of the corresponding image source by the microphone array. It is evident the importance that the method can be used under conditions of partial visibility.

The Figure 7.15 shows a detail of the area of interest for the acquisitions made. This is the region actually reconstructed in the anechoic chamber as the remaining walls have negligible effects on the acquired signal. In Figure 7.13, we show the experimental setup for the configuration.

As previously mentioned we use a linear and uniform array of 16 microphones to accomplish these acquisitions. The positions of the direct sources are shown in Figure 7.15 with red circles. The dotted red lines are the estimated walls.

To perform the estimations for the straight lines of visible reflectors we use a mean of the estimated straight lines for all the direct sources that allow

Figure 7.13: Experimental setup for the last configuration.

to see the corresponding image sources.

This experiment shows the effectiveness of the proposed method in real scenarios for more complex environments.

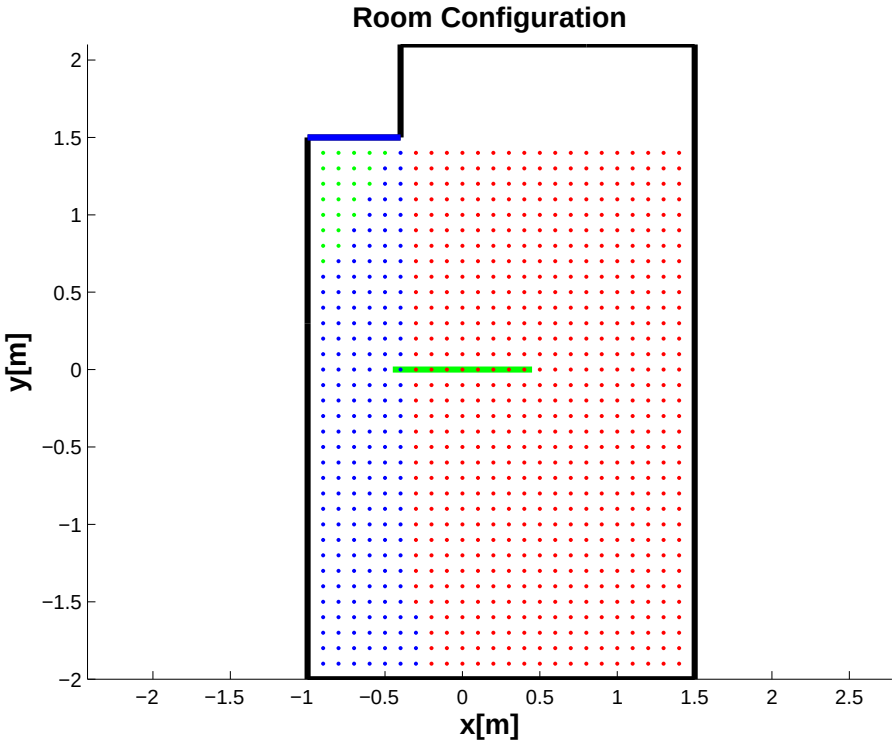


Figure 7.14: Configuration of the simulated room in the experiment. It is pointed out the visibility of the blue wall by the green microphone array dividing the room in completely visible (green points), partially visible (blue points) and non-visible (red points)

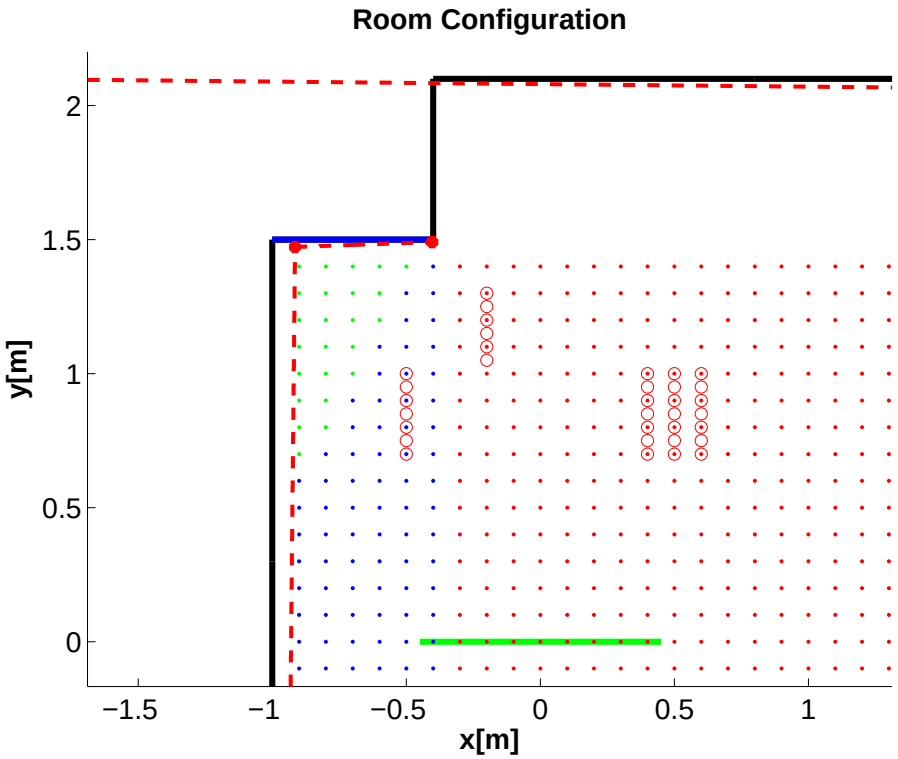


Figure 7.15: The detail of the configuration used for the experiment with the extimated walls(red dotted lines) and endpoints (red dots). The red circles are the positions of the used direct sources.

Conclusions

This work describes an innovative approach to the problem of inference on the geometric and acoustic properties of the environment through the simultaneous use of sensors and sound sources. The information resulting from the processing of the acquired signals are mapped into a space called RaySpace which corresponds to a sampled and distorted version of the Plenacoustic function of the environment. The RaySpace parametrization allows to define the concept of acoustic images. The acoustic images are an overall description of the acoustic scene and contain information about the geometric and radiometric properties of the environment.

The thesis has shown that such acoustic images can be used in an appropriate way to estimate the positions of the reflective surfaces. Furthermore, the new approach to the problem shows interesting properties with respect to the methods already reported in the literature. In particular it allows to estimate not only the line corresponding to the interested reflector but also its size and its extreme points.

Although in this work the obtained acoustic images are used to infer only on the geometric properties, the information in the acoustic images go beyond the geometry of the environment. These snapshots of the acoustic scene also contain numerous other information such as the reflection coefficients of the walls and the radiation pattern of the sources. Consequently, application of the techniques described in this paper will lead to further progress within the environment-aware processing. The potential of the method are not limited to the estimation of the acoustical properties of the environment but may lead to develop innovative methods that exploit this information. For example it may allow to predict the acoustic pressure field in generic positions in space through a finite and spatially limited acquisition.

Bibliography

[1] European project SCENIC, Self-Configuring ENvironment-aware Intelligent aCoustic sensing. `http://www-dsp.elet.polimi.it/ispg/SCENIC/`.

[2] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio. L1 regularized room modeling with compact microphone arrays. *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)), Dallas, Tx*, pages 157–160, Mar 2010.

[3] T. Betlehem and T.D. Abhayapala. Theory and design of sound field reproduction in reverberant rooms. *Journal of the Acoustical Society of America*, 117:2100–2111, 2005.

[4] F. Alton Everest, Ken C. Pohlmann. *Master Handbook of Acoustics (Fifth Edition)*. McGraw Hill, 2009.

[5] Ezio Zandegiacomo. Elementi di acustica.

[6] Jerry C. Whitaker. *Master Handbook of Audio Production*. McGraw Hill, 2003.

[7] Mark Kahrs, Karlheinz Brandenburg. *Applications of digital signal processing to audio and acoustics*. Kluwer Academic Publishers, 2002.

[8] John Eargle. *The Microphone Book (Second Edition)*. Focal Press, 2004.

[9] 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.

[10] S. Tervo and T. Korhonen. Estimation of reflective surfaces from continuous signals. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 153 –156, march 2010.

[11] 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.

[12] 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.

[13] 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.

[14] 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*, pages 17 –20, oct. 2011.

[15] 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.

[16] 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*, pages 2620 –2623, may 2011.

[17] 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*, pages 153 –156, may 2011.

[18] 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*, pages 113 –116, may 2011.

[19] 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*, pages 321 –324, may 2011.

[20] 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*, PP(99):1, 2012.

[21] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[22] J. Berent and P.L. Dragotti. Plenoptic manifolds. *Signal Processing Magazine, IEEE*, 24(6):34 –44, nov. 2007.

[23] 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*, SIGGRAPH '96, pages 43–54, New York, NY, USA, 1996. ACM.

[24] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, pages 31–42, New York, NY, USA, 1996. ACM.

[25] T. Ajdler, L. Sbaiz, and M. Vetterli. The plenacoustic function and its sampling. *Signal Processing, IEEE Transactions on*, 54(10):3790 –3804, oct. 2006.

[26] D. Markovic, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro. Visibility-based beam tracing for soundfield rendering. In *Multime-*

dia Signal Processing (MMSP), 2010 IEEE International Workshop on, pages 40 –45, oct. 2010.

[27] M.R. Azimi-Sadjadi, A. Pezeshkib, L. Scharfb, and M. Hohil. Wideband doa estimation algorithms for multiple target detection adn tracking using unattended acoustic sensors. *Proc. of Spie- The international Society For Optical Engineering*, 2004.

[28] P. Stoica and A. Nehorai. Music, maximum likelihood, and cramer-rao bound. *IEEE transactions on acoustics, speech and signal processing*, 1989.

[29] D. Feldman. Algorithms for finding the optimal k-line-means. Master's thesis, Tel-Aviv university, 2003.