Politecnico di Milano
Facoltà di Ingegneria dell'Informazione

Polo regionale di Como
Master of science in computer engineering

Environment Estimation from Acoustic Measurements: a Geometric Approach

Supervisor: Prof. Augusto Sarti
Assistant Supervisor: Dott. Fabio Antonacci

Master graduation thesis by:
Erica Nastasia
Student Id.number 724646

Academic year 2009-2010

Politecnico di Milano
Facoltà di Ingegneria dell'Informazione

Polo regionale di Como
Corso di laurea specialistica in Ingegneria Informatica

Ricostruzione dell'Ambiente da Misure Acustiche: un Approccio Geometrico

Relatore: Prof. Augusto Sarti
Correlatore: Dott. Fabio Antonacci

Tesi di Laurea Specialistica di:
Erica Nastasia
Matricola 724646

Anno Accademico 2009-2010

Sommario

Questa tesi affronta il problema della ricostruzione della geometria di un ambiente a partire da emissioni sonore e dalle acquisizioni delle relative risposte ambientali. Lo scenario applicativo in cui questo problema si potrebbe presentare è quello di un'applicazione avanzata di analisi e/o sintesi audio, che trarrebbe vantaggio dalla capacità di percepire l'ambiente in cui è immersa. La soluzione proposta in questa tesi sfrutta più microfoni omnidirezionali e altoparlanti direttivi, disposti casualmente o secondo una schiera. Si descrive dettagliatamente come ricavare dalle misure effettuate una sorta di mappa delle riflessioni ambientali, e successivamente come inferire la geometria dell'ambiente. Il metodo è stato validato con simulazioni ed esperimenti.

Abstract

This thesis deals with the problem about reconstructing the geometry of an environment through emissions and respective acquisitions of sound responses. The application scenario in which we can find this problem, is an advanced analysis/rendering audio application, which would take advantage of the ability to infer the environment in which it is immersed. The proposed solution in this thesis exploits multiple omnidirectional microphones and directive sources, placed randomly or depending on an array. We describe in detail how it is possible obtain from measures a kind of map of reflections, and then how we can infer the geometry of the environment. The method has been validated by simulations and experiments.

Ringraziamenti

I ringraziamenti vanno innanzitutto a Fabio, per avermi aiutato e sostenuto nello svolgimento del lavoro presentato e nella stesura di questa tesi. Lo ringrazio inoltre per la pazienza con cui ha più volte riletto e corretto il presente lavoro. Ringrazio inoltre il prof. Augusto Sarti, per la fiducia e l'opportunità offerta. Ma i più grandi ringraziamenti sono rivolti alla mia famiglia, alla quale devo tutto, quello che sono diventata e quello che potrò essere. Se non fosse per i loro sforzi non sarei qui oggi e non esisterebbe questa tesi. Un grazie più particolare va al mio fidanzato, che mi ha supportato moralmente in tutti questi mesi.

Contents

List of Figures

List of Tables

Notation

- $\boldsymbol{v}$ vector
- $|\boldsymbol{v}|$ euclidian distance(norm)
- $\hat{v}$ unitary vector directed as $\boldsymbol{v}$
- $x(t) \otimes y(t)$ cross-correlation: $\int_0^\infty x(t)y(t+\tau)\,d\tau$

Chapter 1

Introduction

The role of the environment in rendering and sound synthesis applications is essential, since it greatly influences the propagation of sound. Echoes, reflections and reverberations can affect the performance of a rendering system, whose effect may significantly differ from the predicted one. Moreover, the environment influence is not easy to detect and isolate. This is most frequent in complex applications that make use of microphone and loudspeaker arrays, like Wavefield Synthesis, dereverberation, etc. One way to mitigate the environmental influence is the use of muffling panels in rendering applications, or to design more robust algorithms for analysis tasks.

Nonetheless, the environment has not exclusively a negative connotation. On the contrary it can be a source of useful information: many space-time algorithms could greatly benefit from the knowledge of the geometry of the environment. The involved algorithms can be projected as robust as possible also thanks to the considerable hand that the knowledge of the environment in which the application lies can provide.

These are the main reasons for which the issue of extracting information from the environment and how to use it, is the subject of several actual projects. Among these we mention SCENIC (Self-Configuring Intelligent Environment-aware acoustic sensing,[29]), European project which this thesis aims to contribute. This contribution concerns the problem of creating an analysis /rendering sound application, which has the ability to infer the geometry of the environment in which it is immersed.

In this work of thesis we deal with the problem of reconstruction of the geometry of the environment by using acoustic measurements. The method makes use of multiple microphones and loudspeakers, in order to study the

response of the environment from various points, and an extension to multiple reflectors is also presented.

In the first section we propose a general introduction to address the problem of reconstruction of the scene. Our method is outlined in more detail in the second section of this chapter. Finally, in the third section, we provide a scheme of the rest of the thesis.

1.1 The acoustic scene reconstruction

The acoustic reconstruction of the geometry of the environment aims to obtain a representation of 3D environment through emissions of a sound with a prescribed space-time structure. The approach is to derive the model that generated the data from the data themselves, so we classify the problem, by an algorithmic point of view in the category of inverse problems. Therefore, in our case, acoustic measurements represent data and 3D scene refers to the model. A drawback coming from the category of inverse problems, is that they are generally ill conditioned, thus they typically rely on strong assumptions on the model and require very complex algorithms.

Solutions existing in literature for reconstructing the geometry of the environment are many and based on various methodologies. One way to address the problem of reconstruction of the scene is the optical one, which actually represents the most mature approach, and which is based on images captured by cameras. Typically two or more cameras capture the scene from different points of view, and information is extracted from 2D images.

Another approach is the scene reconstruction by the acoustic method. It is based on the emission of sound signal and acquisition of related backscattered signal by a set of microphones arranged in an array or randomly. The range information can be obtained directly from the estimate of the time of flight of the signal. Its application field has historically been limited to underwater navigation and mapping, where the turbidity of water precludes the use of cameras.

In this thesis, we adopt an acoustic approach. In fact, optical methods would introduce too many unnecessary details, since we are interested to obtain an approximate estimate of the geometry of an environment. Moreover, optical approaches can also turn out to be harmful: they can estimate reflective walls, even if they are completely muffling obstacles.

Furthermore, our thesis starts from a method for reconstructing the 2D

geometry of the environment, in order to provide a formulation and a solution of the problem also in the 3D situation. The approach in the two-dimensional case is based on the signals acquired by microphones, when a series of acoustic stimula are produced by loudspeakers, in different positions in space. After estimating the times of flight of the reflective paths, we turn them into projective geometric constraints, used for inferring the locations of the reflectors. We analyze the method more in detail in next sections.

1.2 Overview of the solution

The goal, in the process of reconstruction of the scene, is to estimate the geometry of the environment, i.e. a room. The optical methods provide good results, but, as already said, we choose a sound approach. In order to address this problem we use multiple microphones and multiple sources. Microphones can be arranged in various configurations. The reconstruction algorithm is based on the acquisition of acoustic measurements considering a source that emits a signal and source-microphone pairs for all microphones inside the environment. A similar result would be obtained considering all possible sources in relation to a specific microphone, thanks to the *reciprocity theorem*. Using multiple sources increases the robustness of the algorithm, since as the number of recorded acquisitions becomes larger as more data are used to give confirmations of correctness of results. However in all cases a reasonable estimate of the geometry of the environment can also be deduced using only one speaker and a suitable number of microphones.

Regarding the reconstruction of the scene, we have to define a model for the geometry of the environment and one for the propagation of sound. About the first model we assume that the environment consists of planar faces, which properly approximate a room. The second model instead refers to optical acoustics, whereby the sound waves propagate along rays, as light does. This model is acceptable in typical frequency range of the audio hardware we chose for a medium sized environment. We will see that under these assumptions, the influence of the environment is to create a set of virtual images of the active sources, known as *image sources*. As far as the model is valid this is the only physical effect that takes place, thus it is the only thing that can be measured. Therefore, the reconstruction of the scene can be divided in two steps: it is initially based on the calculation of the position of the image sources created by all the reflectors in the scene, and later

on creating an acoustic map of the environment, that describes the distribution of reflections in the room and thanks to which it is then possible to infer its geometry. In this second task the main problem is represented by multiple reflections: in addition to the direct reflections the system detects also higher order reflections, which are not meaningful by themselves and are not easily distinguished from the first order ones. It would be quite easy to integrate the method to provide more informative data (employing more microphones), but we chose to stick to a minimalist approach, to show that scene reconstruction is possible even from such poor data. However, an in-depth analysis of this problem is beyond the scope of this thesis. We confine ourselves to take in account for computation only first order reflections.

1.3 Outline of thesis

An important factor in designing an approach for the reconstruction of the scene is the model of propagation of the medium. The model we assume requires a detailed description of the most important laws on which it is based on. This part is addressed in Chapter 2 of this thesis.

In Chapter 3 we propose a brief overview on the state of the art in the reconstruction of the scene, both in acoustic field and optical one, to give an idea of other possible solutions for similar problems.

Chapter 4 describes how the measures obtained from the times of reflection can be turned into constraints acting directly on geometric primitives, that is sources, receivers and reflectors. More in detail, each measurement on acoustic paths generates a projective constraint that is well represented by a quadratic form. Thus, we investigate better how the transition from measures to constraints occurs and we analyze in detail the specific quadric suitable to represent constraints in a three-dimensional situation. In particular, we mention the projective geometry as an approach to the study of these specific constraints and we propose two methods to derive the equation that represents the involved quadratic form.

In Chapter 5 we analyze the approach used in order to infer a reflector from multiple measures and constraints, exploiting the Hough transform as a method of resolution for this issue. Moreover, we show how, starting from geometric constraints, we obtain a system of equations, each of them corresponding to a single constraint, and how the approach turns into adopting

a non-linear Least Squares formulation of the problem. Therefore, the next step includes the minimization of the cost function resulting from the system, where the computed minimum represents the estimate of the reconstruction process. Furthermore, since the task of minimization detects several local minima, we also discuss how the Hough transform can be used in order to obtain one global minimum.

Chapter 6 describes how the use of Hough transform can be extended to allocate and group various reflections by the reflector from which they have been generated; therefore we analyze how the method is extensible to multiple reflectors to infer then the geometry of the environment.

Various simulations and experiments are presented in Chapter 7, which validates techniques we have proposed.

In Chapter 8 we propose some future directions, providing a starting point for more elaborate works, particulary about the diffraction theory. We also show a brief mathematical analysis of the issue.

Finally, in Chapter 9 we draw the conclusions of this thesis.

Chapter 2

Background

If we want to use the sound as means for sensing and knowing the environment we have to be able to predict the interactions of sound waves with the objects in the scene. For this aim, the knowledge of concepts at the base of the theory of propagation assumes crucial importance. The model we adopt is pretty simple, and it is the *optical acoustics*. It assumes that sound waves propagate along rays, as for the light. Although this is a rough approximation, it is acceptable in the frequency range of hearing and sound in a medium sized room. We know from the theory of illumination, that an area hit by a beam of light can behave as a diffuse surface, distributing homogenously light in all directions, or as a specular surface, perfectly reflecting the beam of light, or it can give rise to diffraction phenomena. In our case, we suppose that sound rays, are reflected specularly from hit obstacles, neglecting diffusion and diffraction phenomena (actually, the case of obstacles that diffract sound will be widely discussed in Chapter 8).

In the first section we derive the wave equation. In the second one we analyze the acoustic optics. In the third section we mention the reciprocity theorem, which will turn to be useful in the next discussion. Finally, in the fourth section, we introduce the main mathematical notions belonging to the *projective geometry*.

2.1 Acoustic wave equation

In this section we derive the acoustic wave equation, which will be the basis for all our future developments.

2.1.1 The general case

This paragraph is dedicated to infer the more general equation for acoustic waves.

Consider an infinitesimal volume element of air, centered in $\boldsymbol{r}$, with ρ the density of the volume element, P the pressure exerted on it and $\boldsymbol{u}$ its speed. Remembering also that the symbol $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ represents the 3-dimensional Laplacian operator, the *continuity equation* and that of *conservation of momentum* in three dimensions for lossless fluid can be written respectively as

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \boldsymbol{u}) = 0,$$

and

$$\rho[\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}] + \nabla P = 0,$$

where

$$\rho = \rho_0 + \delta\rho, \quad |\delta\rho| \ll \rho_0,$$

and

$$P = p_0 + p, \quad |p| \ll p_0,$$

with $\rho_0 = 1.18\frac{Kg}{m^2}$ is the air density and $p_0 = 101300Pa$ the atmospheric pressure, both in standard conditions. The corresponding linear forms are in fact described respectively by

$$\frac{\partial}{\partial t}(\delta\rho) + \rho_0(\nabla \boldsymbol{u}) = 0, \tag{2.1}$$

and

$$\rho_0 \frac{\partial \boldsymbol{u}}{\partial t} + \nabla p = 0. \tag{2.2}$$

In order to obtain the acoustic wave equation, we need a third equation in addition to (2.1) and (2.2). For this aim we briefly describe where it is derived from.

For any fluid (liquid or gas) that moves inside a duct with section S we can write the general isentropic equation of state as a Taylor series. We obtain

$$P = p_0 + A\frac{\rho - \rho_0}{\rho_0} + \frac{B}{2!}\left(\frac{\rho - \rho_0}{\rho_0}\right)^2 + \cdots, \tag{2.3}$$

Let's introduce the sound speed c, defined by

$$\begin{aligned} c^2 &= \left.\frac{\partial P}{\partial \rho}\right|_{S = \text{ constant}} \\ &= \frac{dP}{d\rho} \quad \text{(for an isentropic process).} \end{aligned}$$

After some calculations in (2.3) and approximating $\rho \to \rho_0$, c^2 becomes a constant, denoted by c_0^2, or *small-signal sound speed*, so we can write

$$p = c_0^2 \delta\rho. \tag{2.4}$$

Putting together equations (2.1), (2.2) and (2.4) we obtain the ***acoustic wave equation***

$$\nabla^2 p(\boldsymbol{r}, t) - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}(\boldsymbol{r}, t) = 0. \tag{2.5}$$

Furthermore, if the fluid is a gas, knowing for an adiabatic transformation we state that $\frac{P}{p_0} = \left(\frac{\rho}{\rho_0}\right)^\gamma$, and linearizing as before, through simple calculations we find out that (2.5) can be rewritten as

$$\nabla^2 p(\boldsymbol{r}, t) - \frac{\rho_0}{\gamma p_0}\frac{\partial^2 p}{\partial t^2}(\boldsymbol{r}, t) = 0,$$

where $c_0 = \sqrt{\frac{\gamma p_0}{\rho_0}}$ is the sound speed propagation, which in standard conditions is $343\frac{m}{s}$.

If we consider planar waves, in which pressure varies only along one direction, like x axis, the solution of (2.5) has the form

$$p(x, t) = f(t - \frac{x}{c}) + g(t + \frac{x}{c}).$$

Functions f and g are arbitrary (the only condition is that they be twice integrable). Usually, harmonic solutions are preferred, since they appear to be efficient for many acoustic phenomena, and are easy to handle (through the Fourier transform).

2.1.2 The spherical case

In this paragraph we extract the acoustic wave equation for spherical waves. Generally, dealing with non planar waves, it is useful to introduce a new variable, the *velocity potential* Φ, defined by

$$\boldsymbol{u} = \nabla\Phi. \tag{2.6}$$

Substituting in (2.2) we obtain

$$p = -\rho_0 \frac{\partial \Phi}{\partial t}. \tag{2.7}$$

Putting together equations (2.1), (2.6), (2.7) and (2.4), we can also write the general acoustic wave equation defined in the previous section, as

$$\nabla^2 \Phi - \frac{1}{c_0^2} \frac{\partial^2 \Phi}{\partial t^2} = 0. \tag{2.8}$$

As before we can write the one-dimensional form of (2.8), not dependent by angular coordinates, as

$$\frac{\partial^2 (r\Phi)}{\partial r^2} - \frac{1}{c_0^2} \frac{\partial^2 (r\Phi)}{\partial t^2} = 0,$$

or

$$\Phi = \frac{f(r - c_0 t)}{r} + \frac{g(r + c_0 t)}{r}, \tag{2.9}$$

where $r = \sqrt{x^2 + y^2 + z^2}$.

We can observe that the equiphase and equiamplitude surfaces are spherical, with center in the origin of the reference frame. Notice also that the wave amplitude is inversely proportional to r (the distance in which we evaluate the sound field).

Consider now, a system of spherical coordinates with the addition of angular coordinates, such that

$$x = r \sin\theta \cos\phi, \qquad y = r \sin\theta \sin\phi, \qquad z = r \cos\theta,$$

equation (2.8) becomes

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial \Phi}{\partial r}\right) + \frac{1}{sin\theta}\frac{\partial}{\partial \theta}\left(sin\theta \frac{\partial \Phi}{\partial \theta}\right) + \frac{1}{sin^2\theta}\frac{\partial^2 \Phi}{\partial \psi^2}\right] - \frac{1}{c_0^2}\frac{\partial^2 \Phi}{\partial t^2} = 0.$$

The equation can be solved through the method of separation of variables, assuming a solution like

$$\Phi = R(r)\Theta(\theta)\Psi(\psi)T(t).$$

The obtained solution is

$$\Phi(r, \theta, \psi, t) = \begin{Bmatrix} j_n(kr) \\ \eta_n(kr) \end{Bmatrix} \begin{Bmatrix} \cos(m\psi) \\ \sin(m\psi) \end{Bmatrix} P_n^m \cos(\theta) \begin{Bmatrix} \cos(\omega t) \\ \sin(\omega t) \end{Bmatrix}, \tag{2.10}$$

where $j_n(kr)$ and $\eta_n(kr)$ are respectively the spherical Bessel and Neumann functions, $P_n^m \cos(\theta)$ represent the Legendre polynomials and $k = \frac{\omega}{c}$. This formulation is often used for standing wave problems.

Conversely when we deal with radiation problems, following formulation of (2.10) is preferred

$$\Phi(r, \theta, \psi, t) = \begin{Bmatrix} h_n^{(1)}(kr) \\ h_n^{(2)}(kr) \end{Bmatrix} \begin{Bmatrix} e^{jm\psi} \\ e^{-jm\psi} \end{Bmatrix} P_n^m \cos(\theta) \begin{Bmatrix} e^{j\omega t} \\ e^{-j\omega t} \end{Bmatrix}, \tag{2.11}$$

where $h_n^{(1)}(kr)$ and $h_n^{(2)}(kr)$ are the spherical Hankel functions, and more precisely $h_0^{(1)}(x) = -j\frac{e^{jx}}{x}$ and $h_0^{(2)}(x) = j\frac{e^{-jx}}{x}$.

Furthermore, we are able to derive a one-dimensional form of the acoustic wave equation also through (2.10) and (2.11); we are interested to do it because we want to show another more detailed formulation of the equation about radiation problems.

In fact, it is easy to see that, by using (2.11) as a general solution since the sound field we are considering is progressive, and not standing and dealing with outgoing spherical waves, we have to choose the combination

$$\Phi_{mn} = \begin{Bmatrix} \cos(m\psi) \\ \sin(m\psi) \end{Bmatrix} P_n^m \cos(\theta) e^{j\omega t} h_n^{(2)}(kr).$$

For the most part of sources in practice, sound field is symmetric with respect to the main propagation axis. If this axis is chosen as the z axis, the solution will not depend on the axial angle ψ. In mathematical terms, this condition is obtained taking $m = 0$. Therefore, the more general solution becomes

$$\Phi = \sum_{n=0}^{\infty} A_n P_n \cos(\theta) e^{j\omega t} h_n^{(2)}(kr). \tag{2.12}$$

Moreover, if the source emits sound equally distributed in each direction, the sound field does not depend on angle θ. But since the only one term for which the sum would not depend on θ is the zero-order Legendre polynomial ($P_0 \cos(\theta) = 1$), all coefficients A_n except A_0 must be zero. Equation (2.12) becomes

$$\Phi = \Phi_0 = A_0 e^{j\omega t} h_n^{(2)}(kr),$$

i.e.

$$\Phi = A_0 j \frac{e^{j(\omega t - kr)}}{kr},$$

that represents the field generated by a pulsating sphere. In this case we speak of *monopole* because the sound is radiated homogeneously in all directions. In practice, it is important to notice that both the wave amplitude and phase depend on distance: the first one decreases as r increases and similarly the time of arrival t referred to direct path of the wave, i.e. the time spent since it was generated, depends on position.

Actually it is another form of (2.9), but while this is used especially treating with enclosure problems, that extracted now is preferred about radiation problems.

2.2 Optical acoustics

In this section we introduce the concept of *acoustic ray*, analyzing under which assumptions we can define it. Moreover, we address the problem of how the rays behave, during the propagation, in the interaction with an obstacle; in order to do this, we describe principles on which *optical acoustics* theory is based.

2.2.1 Acoustic rays and travel time

In this paragraph we analyze the assumptions under which we define an *acoustic ray*.

For our purpose we have to introduce the *eikonal equation*, which provides a link between physical (wave) optics and geometric (ray) optics. We consider now the general case in which the propagation speed c depends on position.

If we consider the following solution of the pressure field as

$$\Phi = A(\boldsymbol{r})e^{j(k\boldsymbol{r}-\omega t)} = A(\boldsymbol{r})e^{-j\omega T(\boldsymbol{r})},$$

where $k = \frac{\omega}{c}$, while $T(\boldsymbol{r})$ is denoted with travel time and it is equivalent to a phase.

In order to use the acoustic wave equation we need some calculations before. First, we have to compute $\nabla\Phi$

$$\nabla\Phi = \nabla A e^{-j\omega t},$$

thus

$$\nabla^2\Phi = \left[\nabla^2 A - \omega^2 A|\nabla T|^2 - j(2\omega\nabla A\nabla T + \omega A^2 T)\right] e^{-j\omega t}.$$

Secondly we have to compute the derivative of Φ respect with the time

$$\frac{\partial^2 \Phi}{\partial t^2} = -A\omega^2 e^{-j\omega t}.$$

Therefore, knowing that

$$\nabla^2 \Phi - \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = 0,$$

we can write

$$\nabla^2 A - \omega^2 A|\nabla T|^2 - j\omega(2\nabla A \nabla T + A^2 T) = \frac{-A\omega^2}{c^2}.$$

The imaginary part gives us information about the amplitude, while the real one provides information about propagation. So we look only the real part and state that

$$\nabla^2 A - \omega^2 A|\nabla T|^2 = \frac{-A\omega^2}{c^2},$$

from which

$$|\nabla T|^2 - \frac{1}{c^2} = \frac{\nabla^2 A}{A\omega^2}.$$

For ω sufficiently large, $\omega \to \infty$ the equation becomes

$$|\nabla T|^2 = \frac{1}{c^2},$$

i.e.

$$|\nabla T(\boldsymbol{r})| = \frac{1}{c(\boldsymbol{r})}.$$

Finally we find

$$\nabla T(\boldsymbol{r}) = \frac{1}{c(\boldsymbol{r})}\hat{k} = |s|\hat{k} = \boldsymbol{s},$$

where $\boldsymbol{s}$ is the *slowness vector*.

It finds out that gradient of a wavefront at a position $\boldsymbol{r}$, here defined as the travel time i.e surface of equal phase, is equal to the local slowness. *The direction of maximum change of the wavefront defines the direction of the wave propagation.* There are several implications. One is that *acoustic rays are perpendicular to wavefronts.* In addition to this, the slowness gives the gradient of the travel time, which specifies the direction of the ray. Each time $c(\boldsymbol{r})$ changes, the gradient of T has to change, and the direction of propagation changes at the same time.

If $c(\boldsymbol{r})$ is known, there is a way to reconstruct the direction of the ray (eikonal ray tracer).

Moreover the eikonal equation is valid simplification for the wave equation under some hypothesis. In the first place ω has to be sufficiently large, but not infinite. Then, change in wave speed along the ray has to be small, i.e. the distance over which $c(\boldsymbol{r})$ changes has to be large compared to the wavelength. Finally, curvature must be small (direction of change in ray path, i.e. ∇T) must be small compared to the wavelength.

2.2.2 Image source method

In this section, we describe how sound field behaves when an acoustic ray interacts with an obstacle.

Assume that the propagation speed is uniform, the reflector has infinite extension and that its influence on the incident wave is independent from direction and the frequency of the wave itself. We analyze the situation starting from planar waves. The *ray* is the normal to the wave front. The **Snell's law of reflection** stases that *the reflected ray forms with the normal to the plane a specular angle with respect to the incident ray*, as shown in Figure 2.1. In practice, it is like if the reflected ray originated behind the reflector.

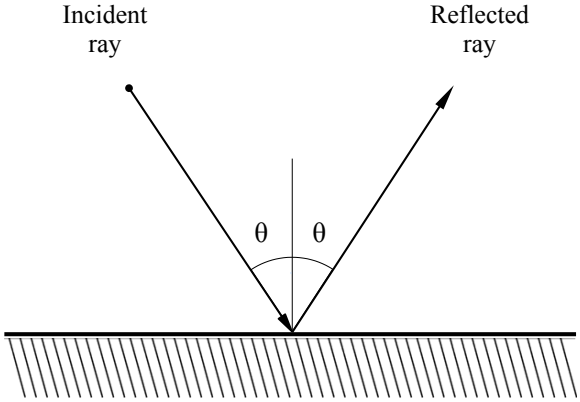


Figure 2.1: Law of reflection. The incident ray and the reflected ray make the same angle with the normal to the plane.

In the case of spherical wave fronts, acoustical optics theory still holds, according to which sound propagation is completely described by acoustic

rays: also in this case we consider the reflected ray as if it were generated by the virtual source whose position is the mirrored image of the real source with respect to the plane. This is the definition of *image source*, and the method that studies the influence of reflectors using image sources is called *image sources method* (Figure 2.2). Since image sources actually behave like real sources, they can in turn give rise to reflections, and so produce second order image sources, third order ones, etc. But in practice, we confine ourselves to the study of the impact of first order reflections, ignoring the contributions from higher orders, for two main reasons. First, high order reflections, are more attenuated than the smaller order ones. Second, the superposition of many reflections may be difficult to analyze with our instruments, but it can instead require a statistical study.

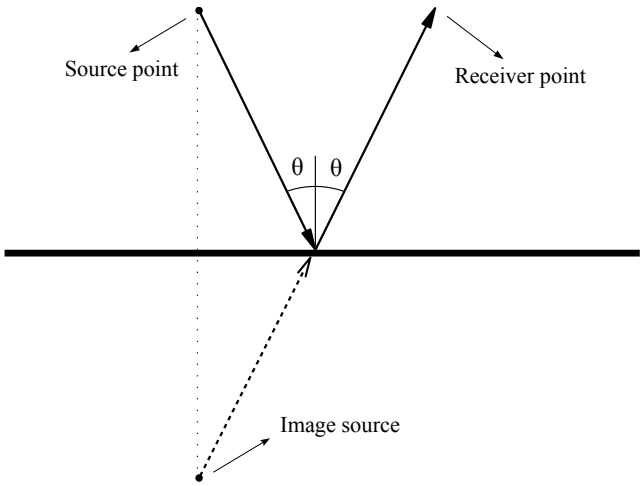
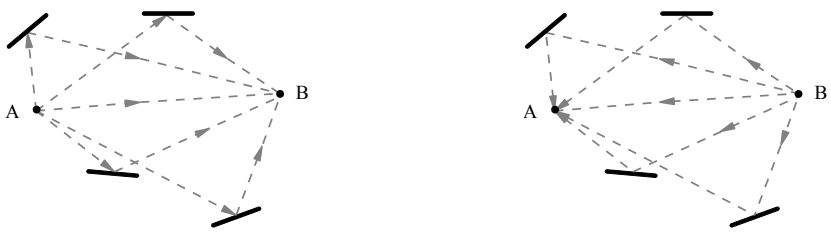


Figure 2.2: Image source method. The reflected ray seems come behind the reflector, from a virtual source that is the specular image of the real one.

2.3 Reciprocity theorem

In this section we explain the reciprocity theorem, which turns out to be useful in many situations. It is a theorem applicable in various fields such as electrostatics, electromagnetism and acoustics. In our case, we will analyze the theorem from the narrow point of view of optical acoustics. In practice, it states that: *the effect of a source placed in A on a receiver placed in B is the same that we would see if the source were in B and the receiver in A.*

Indeed, the concept is fairly intuitive and it is shown in Figure 2.3: the listener in B will receive different replicas of the signal generated in A, each path of which, depend on the reflector with which the signal interferes, determining also the path length. Consequently, swapping positions of the receiver and source, nothing changes, because delays and attenuations of the signal depend exclusively on the reflector crossed along the way and the length of the path itself, remembering that we have assumed a model linear propagation, not only for direct signals but also for reflections. In conclusion, the obtained effect, is that to go through paths of the signals in the opposite direction, since interactions with reflectors and path lengths remain the same.



(a) *Source in A and listener in B.* **(b)** *Source in B and listener in A.*

Figure 2.3: Reciprocity theorem. The effect of a source in A and a receiver in B is the same as the source were in B and the listener in A.

2.4 Projective geometry in 2D

Due to the projective nature of rays, a viable approach to represent propagation of rays is given by *projective geometry*. In the following paragraphs, we present the main mathematical concepts at the base of the theory.

2.4.1 Introduction

We are familiar with Euclidean geometry and with the fact that it describes our three-dimensional world so well. In Euclidean geometry, the sides of objects have lengths, intersecting lines determine angles between them, and

two lines are said to be parallel if they lie in the same plane and never meet. Moreover, these properties do not change when the Euclidean transformations (translation and rotation) are applied. However, when we consider the imaging process of a camera, it becomes clear that Euclidean geometry is insufficient: lengths and angles are no longer preserved, and parallel lines may intersect.

Euclidean geometry is actually a subset of what is known as projective geometry. Projective geometry models well the imaging process of a camera because it allows a much larger class of transformations than just translations and rotations, a class which includes perspective projections. Of course, the drawback is that fewer measures are preserved, certainly not lengths, angles, or parallelism.

Therefore, projective geometry represents a mathematical framework in which to view computer vision in general, and especially image formation. The main areas of application are those in which image formation and/or invariant descriptions between images are important, such as camera calibration, object recognition, scene reconstruction, image synthesis and the analysis of shadows. Many areas of computer vision have little to do with projective geometry, such as texture analysis, color segmentation, and edge detection. And even in a field such as motion analysis, projective geometry offers little help when the rigidity assumption is lost because the relationship between projection rays in successive images cannot be described by such simple and elegant mathematics.

2.4.2 Mathematical notions

As mentioned in the previous paragraph, the projective geometry represents a powerful means, since it is able to handle in a linear and projective way constraints that otherwise would inherently be non-linear. Hence, by means of the projective geometry, we are able to deal with geometric constraints proper to our analysis.

The use of homogeneous coordinates instead of cartesian ones offers several advantages, like representations of non-linear mappings as linear matrix equations, or the uniform handling of points to infinity, over the perfect duality between points and lines. While the use of projective geometry is common in computer vision, it is a quite novel concept in audio processing.

A point in $\mathbb{R}^2$ (the 2D Euclidian space) is represented by a pair of Carte-

sian coordinates (x_1, x_2), that can be seen as the vector $\boldsymbol{x} = [x_1, x_2]^T$. Moreover, a line is represented by the equation $l_1x_1+l_2x_2+l_3 = 0$. A line is so identified by the vector $\mathbf{l} = [l_1, l_2, l_3]$. It is easy to notice that $l_1x_1 + l_2x_2 + l_3 = 0$ and $(kl_1)x_1 + (kl_2)x_2 + (kl_3) = 0$ represent the same line, i.e. the correspondence between lines and vectors is not one-to-one. An equivalence class of vectors under this equivalence relationship define the *homogeneous vectors*. The set of all equivalence classed of vector in $\mathbb{R}^3$, except $[0, 0, 0]^T$ that does not correspond to any line, forms the *projective space* $\mathbb{P}^2$. A point $\boldsymbol{x} = [x_1, x_2]^T$ lies on a line $\mathbf{l} = [l_1, l_2, l_3]^T$ if and only if

$$l_1x_1 + l_2x_2 + l_3 = 0. \tag{2.13}$$

If we represent the point in homogeneous coordinates as the vector $\mathbf{x} = [x_1, x_2, 1]^T$, (2.13) becomes

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

Therefore, as before, the vectors $[kx_1, kx_2, k]^T \in \mathbb{P}^2$ represent the same point $\boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2$.

By using homogenous coordinates, the intersection point between lines $\mathbf{l}$ and $\mathbf{l}'$ is easily found as cross-product of $\mathbf{l} = [l_1, l_2, l_3]^T$ and $\mathbf{l}' = [l'_1, l'_2, l'_3]^T$:

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ l_1 & l_2 & l_3 \\ l'_1 & l'_2 & l'_3 \end{vmatrix},$$

where $\mathbf{l} \times \mathbf{l}'$ denotes the determinant operator.

Analogously, the line $\mathbf{l}$ joining points $\mathbf{x}$ and $\mathbf{x}'$ is given by the cross-product of $\mathbf{x} = [x_1, x_2, 1]^T$ and $\mathbf{x}' = [x'_1, x'_2, 1]^T$:

$$\mathbf{l} = \mathbf{x} \times \mathbf{x}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & 1 \\ x'_1 & x'_2 & 1 \end{vmatrix}.$$

In this representation an arbitrary homogenous vector $\mathbf{x} = [x_1, x_2, x_3]^T$ represents the point $\boldsymbol{x} = [x_1/x_3, x_2/x_3]^T$ in $\mathbb{R}^2$. Notice that when $x_3 = 0$ the two cartesian coordinates go to infinity. These points, with the last coordinate equal to zero are known as *ideal* points, or *points at infinity*. The set of all ideal points may be written as $[x_1, x_2, 0]^T$. This set lies on a single line, the *line at infinity*, denoted by $\mathbf{l}_\infty = [0, 0, 1]^T$ (verify that inner product between the point and the line is equal to zero).

It is necessary to note the importance of introducing points at infinity and how this concept simplifies the intersection properties of points and lines. Assume that we want to find, in euclidian geometry, the intersection of two parallel lines, $\mathbf{l} = [l_1, l_2, l_3]^T$ and $\mathbf{l}' = [l_1, l_2, l'_3]^T$. Using projective coordinates, the third coordinate of the intersection point obtained by the cross-product of the two lines is zero:

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ l_1 & l_2 & l_3 \\ l_1 & l_2 & l'_3 \end{vmatrix} = (l'_3 - l_3)[l_2, -l_1, 0]^T.$$

This result confirms the idea that parallel lines meet at infinity and allows us to consider parallel lines like non parallel ones. Moreover, the role of points and lines can be interchanged in statements concerning their properties. For example the equation $\mathbf{l}^T\mathbf{x} = 0$ is symmetrical, since $\mathbf{l}^T\mathbf{x} = 0$ implies $\mathbf{x}^T\mathbf{l} = 0$, in which the positions of lines and points are swapped. So a general principle holds, the *duality principle*, and it is also present dealing with the definitions of intersection of two lines and lines passing through two points, in which we can still swap roles of points and lines.

A way of thinking of $\mathbb{P}^2$ is a set of rays in $\mathbb{R}^3$. A set of vectors $[kx_1, kx_2, kx_3]^T$ as k varies, forms a ray through the origin. As well as such a single point in $\mathbb{P}^2$ can be represented by a ray, lines in $\mathbb{P}^2$ are planes passing through the origin. A model for the projective plane is proposed in Figure 2.4. Points and lines may be obtained by intersecting this set of rays and planes by the plane $x_3 = 1$. Moreover, the rays representing ideal points and the plane referred to $\mathbf{l}_\infty = [0, 0, k]^T$, $k = 0$ are parallel to the plane $x_3 = 1$.

After this study, we can notice that one limit of traditional projective geometry is that it does not allow to distinguish between sources and receivers, if we think of them respectively as points rays originate from and sink of rays. Another drawback is that classical projective geometry does not allow to determine in which of two halfspaces separated by a line a given point belongs to. Indeed, it is useful mention some concepts belonging to the branch the *oriented projective geometry*. This field of projective geometry takes in account the signs of a scaling factor k. If the line from $\mathbf{x}$ to $\mathbf{x}'$ is represented by

$$\mathbf{l}' = k[l_1, l_2, l_3]^T,\ k > 0,$$

and the line from $\mathbf{x}'$ to $\mathbf{x}$ is represented by

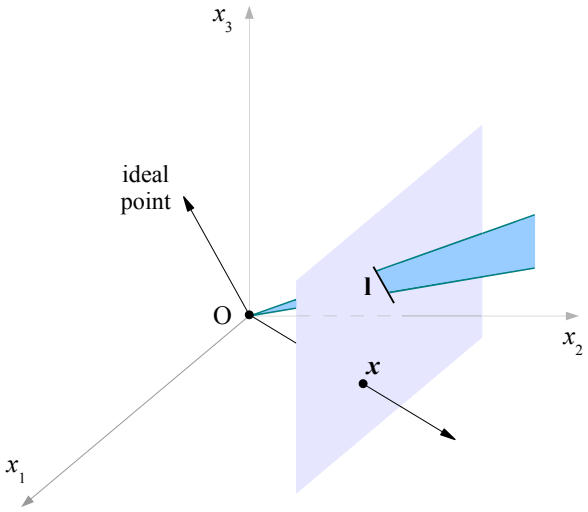


Figure 2.4: Points and lines of $\mathbb{P}^2$ are represented by rays and planes, respectively, through the origin in $\mathbb{R}^3$. Lines lying in the x_1x_2 plane represent ideal points, and the x_1x_2 plane represents $\boldsymbol{l}_\infty$.

$$\mathbf{l}'' = k[l_1, l_2, l_3]^T,\ k < 0,$$

or at the same way

$$\mathbf{l}'' = \mathbf{l}' = -k[l_1, l_2, l_3]^T = k[l'_1, l'_2, l'_3]^T,\ k > 0,$$

the two lines are not coincident thanks to the sign of k.

Therefore, given two points we can find two lines joining them and the not-oriented line is obtained by combining the two oriented lines.

In this representation we are able to distinguish left and right sides of a line. Summarizing, we say that given the line $\mathbf{l} = k[l_1, l_2, l_3]^T$, $k > 0$ and a point $\mathbf{x} = [x_1, x_2, 1]^T$, it lies on the left side of the line if $\mathbf{x}^T\mathbf{l} > 0$ and on its right side if $\mathbf{x}^T\mathbf{l} < 0$ (we neglect the proof of this).

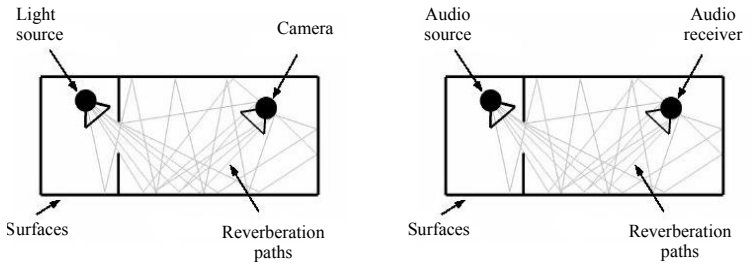
Chapter 3

State of the art in scene reconstruction

In this chapter we offer a survey about the state of the art in the scene reconstruction. We distinguish and describe the optical approach and the acoustical one, used in order to infer geometries.

3.1 Introduction

Sound behaves very similarly to light: the approach that uses reflection paths from a source to a receiver, recalls the art of graphics, that make use of rays of light rather than sound waves, as shown in Figure 3.1. Thus, both approaches model wave propagation. Nonetheless there are some differences between the two methods. One of these regards involved propagation speeds (sound travels more slowly than light) and it causes sound reverberations to



(a) *Light paths.* **(b)** *Acoustic paths.*

Figure 3.1: Comparison between acoustic and light paths of rays.

be perceived over time.

Another difference lies in wavelengths (sound has longer wavelenghts than light) and this affects mainly the size of objects that can be occlusions or create diffraction phenomena. In general, sound diffraction, that arises when an acoustic ray hits an edge or a corner, must be analyzed carefully. Again about sound, specular reflections dominate diffuse reflections and occlusions by small objects do not have considerable effects. Moreover, since sound waves are coherent, modeling phase becomes important.

After this introductory considerations, we present the idea of reconstruction process in the acoustic field as well as in the optical one. There exist many applications for scene reconstruction in both two fields. Optical methods use cameras to acquire a set of images. The reconstruction process usually includes a first step of calibration, i.e. determining cameras parameters (mutual positions, orientations, parameters of lenses, etc). This category of methods can exploit the use of one or multiple cameras, in some cases integrated with mirrors (of several kinds and placed in different ways respect to cameras). Cameras can be fixed or moved, as for mirrors, and sometimes it is possible to capture data with a video by using an handheld camera. In the same way, the model of the scene that algorithms for reconstruction try to fit with acquired data, can considerably vary from one application to another.

Acoustic methods, instead, perform a scene reconstruction on the base of the time of flight of a signal. Also in this category of methods, there are several ways to act. For instance, in underwater applications we use ultrasonic hydrophones as acoustic cameras, but in general, we can find different kinds of sensors (arrays, spherical arrays, etc), used alone, or coupled with suitable optical cameras, in order to obtain a rich information by the fusion of acquired data. In fact, combination of optical and acoustic sensors represents an emerging field of research.

In the modern applications for scene reconstruction, the novelty lies just in the joint use of optical and acoustic techniques, and particulary in the extension of the geometric analysis to the acoustic field. In this regard, various forms of geometric techniques are widely analyzed in SCENIC, project that concerns with inference of the environment through acoustic measurement, of which a deliverable is reported in [29]. Moreover, in *Multiple View Geometry in Computer Vision* ([25]), a whole field as Computer Vision is analyzed, showing how this discipline has strong connections to fields of mathematics and computer science, while those to physics and psychology perception

have been weakened. In particular, the authors cover concepts belonging to the area of *geometric Computer Vision*, that includes descriptions of how the appearance of objects changes with different viewpoints, and how this goal have been achieved just thanks to the use of sophisticated mathematical techniques. The book treats in detail the complex geometric relations between the images of objects, important to analyze, since one of the targets of science is just to provide explanations for appearances. Furthermore, mathematics needed to describe not trivial issues, as geometric concepts and obtained results, or how to express them in algorithmic form so that they can easily be transformed into computer code, over real examples that illustrate the concepts, is exposed in a accessible way.

Summarizing, both optical and acoustic methods count many applications in several fields, thus, altough this specific analysis should not be covered in this work of thesis, we think that is is important to mention some examples of application.

3.2 Optical methods

Optical methods usually are based on the acquisition of two or more images by using cameras from different points of view trying to get a 3D representation mimicking the stereo human vision. This is done by finding a correspondence between points of different images; then, on the basis of the knowledge of the mutual positions of cameras, it is possible to determine 3D coordinates of those points.

Epipolar geometry describes the correlation between the 3D coordinates of the scene and the 2D ones in stereo images. The first step in the scene reconstruction process is the calibration of cameras, i.e. a precise measurement of their mutual positions and orientations, togehter with the parameters of their lenses. Usually this information is obtained by analyzing a series of images acquired for this aim. An example of calibration in proposed in [14], in which it is performed by using images of a known planar pattern shown at a few different orientations. Either the camera or the planar pattern can be moved. Another way to perform calibration is based on images of the scene to be reconstructed, as in [15], in which the technique provides a 3D reconstruction up to a scale factor. In general, these kinds of approaches require either the intrinsic camera parameters, the extrinsic parameters, i.e

the spatial coordinates of at least five world points of that scene, but many more such points are usually required. This specific paper proposes a method using only two camera views, without other camera information available.

Conversely, situations in which just one camera is employed, can be those in which the scene is static. An example is reported in [16], in which the method works with a handheld camera. The intrinsic and extrinsic parameters of the camera are determined at the moment of image capture. Scene reconstruction is obtained from the set of input images by an adaptive space-carving algorithm. A video of the scene is recorded and the successive frames are compared. Then, the scene is reconstructed up to a scale factor. A similar approach can employ a stereo camera, as in [17]. In this case scale information is retained. In fact, crime scene reconstruction is a field that requires high level applications. Sometimes it can happen to integrate cameras with suitable mirrors to increase its effectiveness. In [18], for instance, the authors present a system composed of an array of spherical mirrors and a single high-resolution digital camera oriented toward the array. In this way it is possible to acquire multiple viewpoints through a single shot. Another approach using spherical mirrors is proposed in [19]. The approach is similar to the previous one, since multiple viewpoints are obtained, but in this case the mirror is moved freely within the field of view of a fixed camera. Therefore the method allows us to obtain wide-angle multi-viewpoint images of a wide area. Moreover, other kinds of approaches allow to enlarge the field of view even up to 360°, as described in [20] and [21], in which the effect is obtained employing two cameras directed downwards against two hyperbolic mirrors. In particular in [21] the adopted sensor is constituted by a laser emitter in addition to the camera coupled with the mirrors. The sensor has omnidirectional viewing ability and also provides depth information about the nearby surrounding.

As a conclusion of the proposed survey, we should analyze the algorithms for reconstruction, but this is not covered in our work. In general they aim to fit the model of the scene with the acquired data. There are several kinds of models, more or less specialized. In [22], for instance, it is presented a framework applied to the reconstruction of people. A camera captures simultaneously a moving person from multiple viewpoints and a humanoid model is animated to match the pose at each time frame.

Other techniques include a form of projection of a known pattern on the objects to reconstruct, like illumination. In this way, we are able to predict

how points of space will be illuminated, thus the recostruction of the scene can be performed through a single image. An example of a color structured light technique is presented in [23], in which we can recover object shape by projecting a pattern of stripes of alternating colors and matching the projected color transitions with observed edges in the image.

3.3 Acoustic methods

Acoustic scene reconstruction is based on the computation of the time of flight of a signal, in order to infer the geometry of the environment.

The signal generated by the environment is reflected and reaches the receiver after a time proportional to the distance from the reflector. Therefore, acoustic measurements can reliably provide also information about distance of objects and the main involved directions. Since sources and receivers are little directional, the angular information is obtained using array microphones: the direction of arrival of the signal is obtained by the delay with which it reaches receivers. Human hearing uses essentially this principle for sound localization.

A method for reconstructing the 2D geometry of an environment is presented in [27]. The approach constists of acquiring, by a fixed microhone, signals coming from loudpspeakers placed in different positions in space. After estimating the times of flights of the reflective paths, the acoustic reconstruction of the scene is based on turning each time of arrival into a projective geometric constraint. Through all geometric constraints we can infer the reflectors and their locations.

This thesis aims to discuss the problem of the acoustic recostruction of the scene in the 3D case, extending concepts and solutions coming from the analysis in the 2D situation.

In general, one of the main applications that uses the acoustic method for the scene reconstruction belongs to the field of underwater navigation. In fact, in order to obtain 3D information in underwater applications scene key points from multiple underwater views (either supplied by multiple cameras or by a single moving camera) can be still used to extract 3D estimates. But unfortunately, altough optical approaches provide high resolution and target details, they are constrained by limited visibility range.

Numerous applications for the reconstruction of very large scale scenes,

like sea floors, are proposed in [1]. Also smaller scale applications have been studied, for the purpose of underwater survey and inspection by submersible robots. In [2] and [3] it is discussed how to provide a three-dimensional reconstruction of an underwater environment by using multiple range views from an acoustic camera (an array of ultrasonic hydrophones). The difficulty is to provide a reconstruction on-line with the aim to improve the understanding of a human operator driving an underwater vehicle. The acoustic camera provides a sequence of 3D images in real time, so data must be recorded and fused to generate a 3D mosaic in the form of a mesh, which is rendered through a graphical interface.

Other examples regard probabilistic approaches, as in [4]. The proposed technique allows the coupled reconstruction and restoration of underwater acoustic images, based on the physics of the image-formation process. Beamforming[1] is used to build a range image from backscattered echoes, associated point by point with the information representing the reliability of such an image. Unfortunately, this kind of images is affected by problems due to the nature of the signal and to the related sensing system. For this reason the range and confidence images are modeled as Markov random fields whose associated probability distributions are specified by an energy function, that can model knowledge of the acoustic system, the considered scene, the noise-affecting measures and integrates reliability information. Optimal estimates of the reconstructed range image map and the restored confidence image are obtained by minimizing the energy function using simulated annealing.

Furthermore, combination of optical and acoustic sensors, for instance in underwater 3D acquisition, allows to overcome the shortcomings presented by optical systems. In [5] and [6] the information from an acoustic camera and an optical camera are fused together to obtain a high resolution 3D image of the environment. In this situation the using of an acoustic camera for range measurements is a key advantage, where the associated epipolar geometry of optical and acoustic image correspondences can be described in terms of conic sections. In [7] an opti-acoustic system composed by a single camera and a multibeam sonar is proposed. While cameras become useless beyond a short range, because of turbidity of water, underwater sonars can operate in larger visibility ranges and provide 3D information even in pres-

[1] It represents a technique used in sensor arrays for directional signal transmission or reception. This method is widely applied in acoustic imaging.

ence of water turbidity conditions though at expense of a coarse resolution and harder data extraction. Furthermore, despite the difficulty of combining two modalities that operate at different resolutions, thanks to technology innovations and advances in acoustic sensors, it is possible to generate good quality high-resolution data, useful to design new techniques for underwater scene reconstruction.

On the contrary, outside of the water, optical methods are preferred respect to acoustical ones. In the field of robotics, however, acoustic sensors are used for their low cost and accuracy in estimating the range, also because the angular resolution has less importance for many applications. In fact, this type of sensors, can still provide detailed information on environmental geometry when joined in their use, by appropriate processing techniques. Examples of techniques that are able to reconstruct a 3D scene are proposed in [8] in which a single sensor performs a scanning motion, and in [9] where a sensor that contains two emitting transducers and several receivers is used.

In [10] it is shown how a single transducer is used to estimate the inclination angle of a planar reflector, on the basis of the echo amplitudes only.

All these methods work at frequencies above the audible range. Conversely, in the audio frequencies range, spherical arrays have assumed interest, both for their resolution and for their compactness. Nevertheless, a spherical array appears also as an expensive solution, since it includes many microphones. However, the employing of spherical arrays is widely discussed in [13]. The authors discuss the increasing interest for circular arrays, since they allows to guide a single or multiple beams to any angle in the plane of the array with a desired beampattern. Moreover, it is investigated how one can have control of the vertical spatial response of the array. For this aim an augmented circular microphone array is proposed.

Other examples of adoptions of spherical arrays are presented in [11] and [12]: the authors propose the use of spherical arrays to infer the temporal sequence of reflections in the environment together with their directions of arrival. In particular, in [12] the authors discuss also in which way the combination of microphones and cameras allow the joint audio visual sensing of a scene. The main concept is that both cameras and microphone arrays are geometry sensors, and microphones arrays can be considered as generalized cameras. In fact, as well as cameras that allow to infer the scene geometry togheter with the location and motion of people or other objects, also micro-

phone arrays can provide geometrical information on the source location and allow to separate sound sources in the scene together with the noise suppression. After these considerations, the authors propose a comparison between a general microphone array and spherical microphone arrays, explaining how the latter show a geometry very close to central projection cameras and so how we can effectively apply standard vision based calibration algorithms.

Chapter 4

From measures to constraints

In this chapter we analyze why and how we perform measurements, and we investigate on how turn them into geometric constraints useful for our problem.

In particular, considering the direct path in a measurement, the position of sources and receivers are constrained and this constraint is well represented by a quadratic form. On the other side, with respect the reflected path of the signal, the constraint acts on the plane on which the obstacle responsible of the acoustic path lies. Since the quadratic form is represented by a conic matrix, this last constraint determines on which plane tangent to the conic, the reflector responsible of the acoustic path has to lie. Putting us in the 2D case we talk about *conics*, while in 3D we have to call them *quadrics*. Therefore, we exploit concepts and notions that *projective geometry* offers because it allows us to describe perfectly geometry and constraints we will deal with. In general, a scheme of a geometric analysis approach is like that represented in Figure 4.1: we measure our variables of interest, we convert these measurements into constraints by using projective geometry through which we can estimate our unknown (position of reflector, etc,...), we combine multiple constraints to accomplish our estimate. Variables of interest to measure from acquisitions can be different, as the Time Of Arrival (TOA), the Time Difference Of Arrival (TDOA), the Direction Of Arrival from projective acoustic maps(DOA), etc. In our case, we assume that source is synchronized with receivers, therefore we are able to measure the Times of Arrival. This variable represents a more constrained information with respect to that provided by TDOAs, but at the same time it is more faithful. However, in the following we propose some brief concepts about Time Differences Of Arrival, also

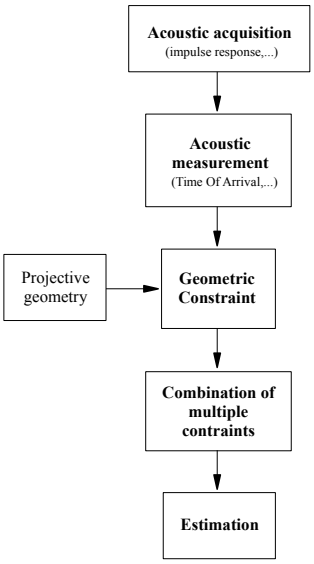


Figure 4.1: General scheme of a geometric analysis approach.

making a comparison between TOAs and TDOAs.

Summarizing, starting from measures of TOAs, we convert them into geometric constraints, that lend themselves to be suitably analyzed by means of projective geometry. Since concepts on which projective geometry is based are not trivial, we start discussing the main principles and the study of constraints in the 2D case. Later, we extend the analysis to the three-dimensional situation, making it easier to understand.

In the first section we provide the definition of Time of Arrival and how measure it. In the second one we shortly describe the Time Differences Of Arrival in relation to Times Of Arrival, focusing on what of similar is retained or not. In the third section we analyze, by means of projective geometry, the geometric constraints derived from TOAs, starting with the treatment in the 2D case for reason of simplicity. In the third section we extend previous concepts to the 3D situation, introducing the conic matrix identifying the quadratic form. Moreover, we analyze two possible methods in order to compute the matrix.

4.1 Time Of Arrival (TOA)

As anticipated before, thanks to the assumption that source is synchronized with receivers, we are able to measure the time of flight of a signal that

propagates from the source to the receivers.

Given the positions of one source and one receiver, we are interested to provide the computation of the Time of Arrival of the acoustic paths that link the two objects, both for direct and reflected signal. We first focus on the case of a single reflector (the extension to multiple reflectors will be treated in Chapter 6).

If the loudspeaker emits a known time-continuous sequence $s(t)$, the signal acquired by the sensor is

$$x(t) = \alpha_d s(t - \tau_d) + \alpha_r s(t - \tau_r) + \nu(t), \tag{4.1}$$

where α_d and α_r are the attenuations of the direct and reflective path respectively, τ_d and τ_r the corresponding times of flight and $\nu(t)$ an additive noise. In order to compute TOAs we need to know the impulse response of the model, and one possible method to do this is the cross-correlation between the signal generated by the source and the signal received by the sensor. Therefore, taking for the source signal a white noise

$$h(t) = x(t) \otimes s(t) = \alpha_d \delta(t - \tau_d) + \alpha_r \delta(t - \tau_r) + r_\nu(t),$$

where $h(t)$ represents the impulse response of the acquired signal, $r_\nu(t)$ the cross-correlation between the signal $s(t)$ and the noise $\nu(t)$, while $\delta(t)$ is the Dirac delta function of the emitted signal.

We observe the presence of two sharp peaks in the impulse response. TOAs are estimated by picking peaks in $h(t)$, i.e. we select the most relevant local maxima. The reflector is responsible of the second of them. An example of impulse response is shown in Figure 4.2. Two significant maxima are visible. The direct path has always the shortest propagation time, thus in the case of a single reflector it is easy to identify the TOA referred to the reflected path. On the contrary, in the situation of multiple reflectors, i.e. multiple peaks together with the direct signal, the observation of a single impulse response is not sufficient to label the time of arrival that identifies the wall that generated it. Therefore, given a value of TOA, in order to know from which reflector has been produced, multiple acquisitions are implied. We will deal with the assignment problem thanks to the joint observation of multiple impulse responses in Chapter 6.

Now, we want to understand how the equations above turn into geometric constraints. This issue is discussed in the third section.

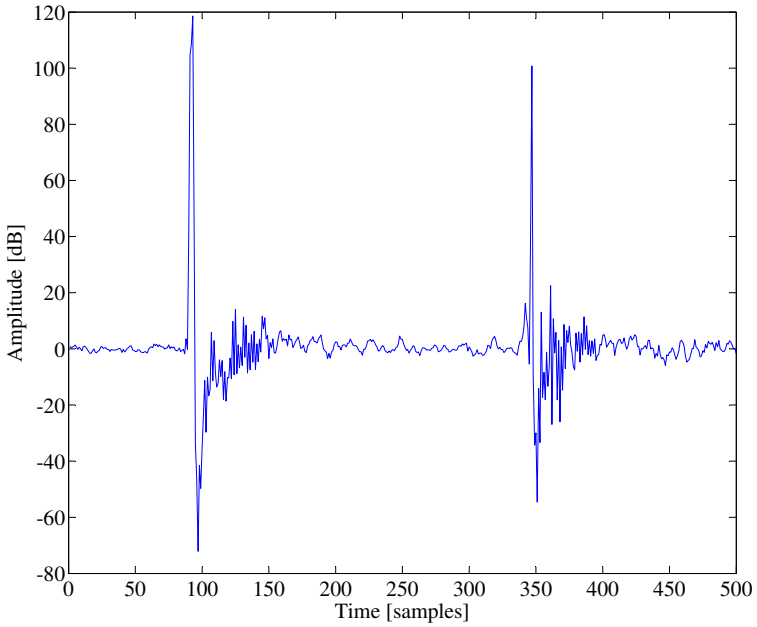


Figure 4.2: Impulse response acquired in anechoic room in which a reflective panel has been placed: the first peak is relative to the direct path, while the second one corresponds to the reflected signal.

4.2 Time Difference Of Arrival (TDOA)

The use of Times Differences Of Arrival as variables of interest in a geometric analysis, represent a subsequent step to the formulation of the Times Of Arrival. Although the TDOAs are consistent with TOAs, we exploit the Times of Arrival as variables of interest for our work, both because of the synchronization between source and receivers, and because of the greater faithfulness of information provided by TOAs, even if the constraints acting on it are many more.

However, in this section we give a short introduction about the formulation of TDOAs, just to underline what are the similarities or differences from the TOAs.

Now, we remove the assumption that the source is synchronized with the receivers, and we retain the synchronism among microphones. This represents those situations in which the source is active in the environment. Thus,

we are not able to compute the Time Of Arrival. However we can obtain information about the source location, by the joint knowledge of the signals at sensors. The signal acquired by the sensor m in the array is

$$x^m(t) = \alpha_d s(t - \tau_d^m) + \alpha_r^m s(t - \tau_r^m) + \nu^m(t). \tag{4.2}$$

The model in (4.2) is exactly equivalent to that in (4.1), except that the source is unknown, along with its delay.

For simplicity we neglect the presence of the reflector. Thus, considering m and k the lags referred to different microphone locations, the cross-correlation between $x^m(t)$ and $x^k(t)$ is

$$R^{mk}(t) = \alpha_d^2 s(t - \tau_d^m) \otimes s(t - \tau_d^k), \tag{4.3}$$

where we assume that the attenuation α_d does not depend on the index of the microphone location and additive noises at different sensors are uncorrelated. As in the case of Times Of Arrival, the approach is to extract peaks from the cross-correlation function.

Different source signals can be used. Considering only the cases of a sinusoidal signal and a white noise, two extreme and different situations arise. We show why this happens. For a sinusoidal signal we obtain

$$R^{mk}(t) = \alpha_d^2 \cos\{2\pi T[t - (\tau_d^m - \tau_d^k)]\},$$

so, maxima are $t = \tau_d^m - \tau_d^k + 2i\pi$, with $i \in \mathbb{Z}$. Moreover, only a limited range of delays is allowed. This is more visible considering a distant source respect to two sensors located in $\boldsymbol{x}^m(t)$ and $\boldsymbol{x}^k(t)$, according to the configuration in Figure 4.3. Thus, the delay in this situation is

$$\tau_d^{mk} = \frac{d}{c} \sin\theta,$$

where d is the distance between the sensors, c the propagation speed and Θ represents the direction from which the plane wavefront comes. Hence, the delay is always in the range $[-\frac{d}{c}, \frac{d}{c}]$ and maxima can be searched in this interval of the cross-correlation function.

Conversely, if the source signal is a white noise, we can write that

$$R^{mk}(t) = \alpha_d^2 \delta[t - (\tau_d^m - \tau_d^k)],$$

with maxima $t = \tau_d^m - \tau_d^k$.

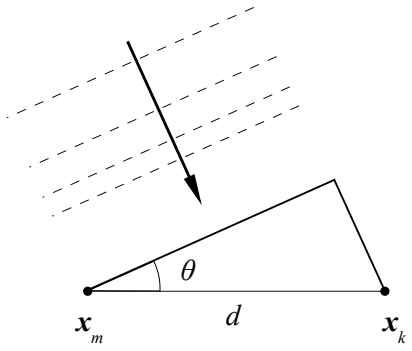


Figure 4.3: The delay between two receivers depends on the distance d between them and the direction from which the wavefront comes.

Summarizing, in the case of a narrowband signal the cross-correlation function is periodic and exhibits multiple local maxima, while with a white noise there is a single maximum. Now, including the presence of one reflector in the treatment, we can rewrite (4.3) as

$$R^{mk}(t) = \alpha_d^2 s(t-\tau_d^m) \otimes s(t-\tau_d^k) + \alpha_d \alpha_r^k s(t-\tau_d^m) \otimes s(t-\tau_r^k) + \alpha_d \alpha_r^m s(t-\tau_d^k) \otimes s(t-\tau_r^m) + \alpha_r^m \alpha_r^k s(t-\tau_r^m) \otimes s(t-\tau_r^k). \quad (4.4)$$

We notice that the function in (4.4) exhibits four local maxima, that is from a single impulse response it is not possible to label a maximum, hence to identify the wall that generates it.

4.3 Constraints related to TOAs (2D case)

In this section we discuss how we can turn measurements of Times of Arrival into geometrical constraints involving sources, microphones and reflectors. For simplicity, we start the explanation with the 2D case, discussing later with the 3D case.

Consider the positions of a source and a microphone, together with one obstacle. Recalling notions acquired in Chapter 2, we can state that the reflector is represented by a line:

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

Consider the loudspeaker placed in S and the microphone in R. The time of Arrival τ_r is the sum of the time of flight from S to P, and from P to R,

where P is the reflection point on the reflector. An illustration is shown in Figure 4.4.

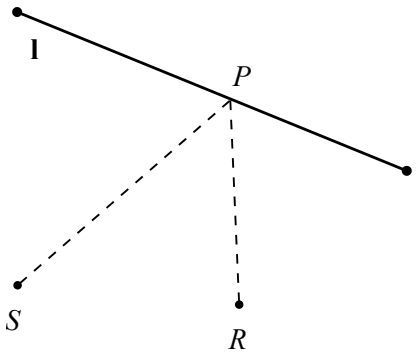


Figure 4.4: The acoustic path of the reflected signal is such that the reflection point P on the reflector $\mathbf{l}$ honors the Snell's law.

In particular, the path length of the reflected signal (first order) is $d_r = c\tau_r$, where c is the sound propagation speed. Therefore we can write:

$$d_r = d_{rS\to P} + d_{rP\to R}. \tag{4.5}$$

Equation (4.5) under which the sum of two distances must be equal to d_r, constrains potential reflection points to lie on an ellipse with foci S and R. Moreover, if we consider the tangent to the ellipse in P, from basic geometry we know that the its perpendicular is also the bisector of the angle $\widehat{SPR}$. It means that every tangent to the ellipse is a potential reflector, as it honors the Snell's law. Therefore, this set of tangent lines to the ellipse can be found by using projective geometry.

We need to define the equation of a conic in $\mathbb{R}^2$:

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f = 0. \tag{4.6}$$

This is a quadratic form, written also in this way:

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

where $\mathbf{x} = [x_1, x_2, 1]^T$ and

$$\mathbf{C} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix},$$

is the *conic matrix*. Note that the conic coefficient matrix is symmetric. Moreover, only the ratios of the matrix elements are important, since multiplying $\mathbf{C}$ by a non-zero scalar does not affect (4.6). Thus, $\mathbf{C}$ is an homogeneous representations of a conic. The conic has five degrees of freedom, which can be thought as the six elements of a symmetric matrix less one for scale.

Summarizing, the set of tangent lines to the conic $\mathbf{C}$, from projective geometry, is

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

where $\mathbf{C}^* = det(\mathbf{C})\mathbf{C}^{-1}$ is the adjoint of the conic matrix $\mathbf{C}$. It is also known as the line conic of $\mathbf{C}$, i.e the bundle of all lines tangent to $\mathbf{C}$.

4.4 Computation of the conic matrix in $\mathbb{R}^3$

In this section we extend our previous considerations to the 3D case, and we show how to compute the conic matrix.[1]

A point $\mathbf{x}$ in $\mathbb{R}^3$ is represented by homogenous coordinates as a 4-vector. In particular the point $\boldsymbol{x} = [x, y, z]^T$, in an homogenous formulation becomes

$$\mathbf{x} = [x_1, x_2, x_3, x_4]^T,$$

with $x_4 \neq 0$, $x = x_1/x_4$, $y = x_2/x_4$, $z = x_3/x_4$.

In $\mathbb{R}^3$ the geometric constraints discussed before still hold, with the addition of one coordinate, ellipses become ellipsoids, still with foci loudspeaker S and receiver R, and lines on which reflectors lie become planes.

A plane in 3D space is described by the equation

$$p_1x + p_2y + p_3z + p_4 = 0.$$

Homogenizing by substituting $x = x_1/x_4$, $y = x_2/x_4$, $z = x_3/x_4$, we obtain

$$p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 = 0,$$

or, in the same way,

$$\mathbf{p}^T\mathbf{x} = 0,$$

[1] In $\mathbb{R}^3$ we have to call it *quadric* matrix.

that means the point $\mathbf{x}$ lies on the plane $\mathbf{p}$. Now we move to the ellipsoid in 3D space. The equation of a quadric in $\mathbb{R}^3$ is:

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 + gx_1 + hx_2 + ix_3 + l = 0. \quad (4.7)$$

The corresponding quadratic form is, as before:

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

but in this case $\mathbf{x} = [x_1, x_2, x_3, 1]^T$ and the conic matrix is

$$\mathbf{C} = \begin{bmatrix} a & b/2 & d/2 & g/2 \\ b/2 & c & e/2 & h/2 \\ d/2 & e/2 & f & i/2 \\ g/2 & h/2 & i/2 & l \end{bmatrix}.$$

We now investigate on how to compute coefficients of the matrix.

4.4.1 Based on the null-space

We can see that the parameter vector obtained by the previous considerations $[a, b, c, d, e, f, g, h, i, l]^T$ is scalable.

In fact the vector $[ka, kb, kc, kd, ke, kf, kg, kh, ki, kl]^T$ refers to the same vector of before. Recalling considerations of the previous section about the conic matrix, we state that in this case degrees of freedom are the ten elements of the symmetric matrix less one for scale. Moreover, from (4.7), each point $\mathbf{x}^{(i)}$ imposes one constraint on the quadric coefficients, since the point must lie on the quadric. Therefore, putting together the constraints from nine points we obtain that the matrix product

$$\begin{bmatrix} x_1^{(1)^2} & x_1^{(1)}x_2^{(1)} & x_2^{(1)^2} & x_1^{(1)}x_3^{(1)} & x_2^{(1)}x_3^{(1)} & x_3^{(1)^2} & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & 1 \\ x_1^{(2)^2} & x_1^{(2)}x_2^{(2)} & x_2^{(2)^2} & x_1^{(2)}x_3^{(2)} & x_2^{(2)}x_3^{(2)} & x_3^{(2)^2} & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & 1 \\ x_1^{(3)^2} & x_1^{(3)}x_2^{(3)} & x_2^{(3)^2} & x_1^{(3)}x_3^{(3)} & x_2^{(3)}x_3^{(3)} & x_3^{(3)^2} & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_1^{(9)^2} & x_1^{(9)}x_2^{(9)} & x_2^{(9)^2} & x_1^{(9)}x_3^{(9)} & x_2^{(9)}x_3^{(9)} & x_3^{(9)^2} & x_1^{(9)} & x_2^{(9)} & x_3^{(9)} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ l \end{bmatrix} = 0.$$

The coefficients of the quadric can be so found thus as the nullspace of the 9×10 matrix. This means that a quadric is determined uniquely (up to scale) by nine points.

In order to compute the null vector, we need to know the nine points the conic is constrained to pass through. This is obtained, by imposing that the path from one focus (loudspeaker) to the point $\mathbf{x}^{(i)}$ lying on the ellipsoid plus the path from latter to the other focus (microphone) is equal to the path length of the reflected signal, as in Figure 4.5 (where for simplicity the microphone is placed in the origin of the reference system). Therefore we

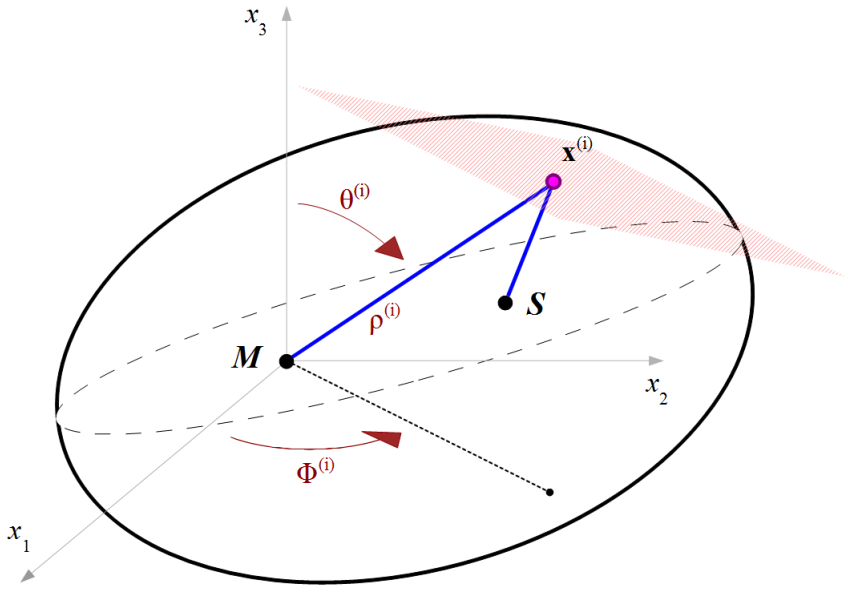


Figure 4.5: Any potential reflector is tangent to the ellipsoid with foci $\boldsymbol{M}$ (microphone) and $\boldsymbol{S}$ (source). The reflected path identified by triangle $\widehat{\boldsymbol{S}\mathbf{x}^{(i)}\boldsymbol{M}}$ honors the Snell's law of reflection.

can write

$$\begin{aligned}\rho^{(i)} + \big[&\left(\rho^{(i)}\sin(\theta^{(i)})\cos(\phi^{(i)}) - x_1^S\right)^2 \\ &+ \left(\rho^{(i)}\sin(\theta^{(i)})\sin(\phi^{(i)}) - x_2^S\right)^2 \\ &+ \left(\rho^{(i)}\cos(\theta^{(i)}) - x_3^S\right)^2\big]^{\frac{1}{2}} = d_r\end{aligned},$$

where $\rho^{(i)}$, $\theta^{(i)}$, $\phi^{(i)}$ are referred to spherical coordinates of the point $\mathbf{x}^{(i)}$, x_1^S, x_2^S, x_3^S are the coordinates of the loudspeaker, while d_r is the path length of

the reflected signal passing through $\mathbf{x}^{(i)}$. More in detail, the first term is the path from the point $\mathbf{x}^{(i)}$ to the microphone, while the second one represents the path from the loudspeaker to the point.

4.4.2 Comparing with a quadratic form

Another way to obtain the parameter vector $[a, b, c, d, e, f, g, h, i, l]^T$ is to use classical geometrical laws for a conic. It is the method we exploit in the rest of the thesis.

For the ellipsoid, particulary, we have that the sum of the distances between foci and a generic point, has to be equal to the double of major semiaxis a. Moreover, consider x_s, y_s, z_s as the coordinates of one focus (source), and x_r, y_r, z_r as those of the other focus (microphone). The Time Of Arrival as we defined it before, is just the time spent to travel from a focus to each other along the reflected path, and this distance is equal to $2a$. In practice, since the TOA is measured in seconds, and the distance in meters, considering c the propagation speed and τ the value of TOA, we write that $2a = c\tau$. For simplicity we assign to $2a$ the variable T, and so we write

$$\sqrt{(x-x_s)^2+(y-y_s)^2+(z-z_s)^2)}+\sqrt{(x-x_r)^2+(y-y_r)^2+(z-z_r^2)}=T.$$

Now, the main approach consists in solving this equation, in such a way that we can compare the expansion with the general form of a quadric given by

$$ax_1^2+bx_1x_2+cx_2^2+dx_1x_3+ex_2x_3+fx_3^2+gx_1+hx_2+ix_3+l=0.$$

In this way we can extract the coefficients of variables. The equation can be solved in implicit form. We neglect calculus and limit to show the result:

$$\begin{aligned}
a &= 4[(x_s-x_r)^2-T^2]\\
b &= 8[(x_s-x_r)(y_s-y_r)]\\
c &= 4[(y_s-y_r)^2-T^2]\\
d &= 8[(x_s-x_r)(z_s-z_r)]\\
e &= 8[(y_s-y_r)(z_s-z_r)]\\
f &= 4[(z_s-z_r)^2-T^2]\\
g &= 4[T^2(x_r+x_s)-(x_s-x_r)(x_s^2-x_r^2+y_s^2+z_s^2-y_r^2-z_r^2)]\\
h &= 4[T^2(y_r+y_s)-(x_s-x_r)(y_s^2-y_r^2+x_s^2+z_s^2-x_r^2-z_r^2)]\\
i &= 4[T^2(z_r+z_s)-(x_s-x_r)(z_s^2-z_r^2+y_s^2+x_s^2-x_r^2-y_r^2)]\\
l &= [(x_s^2+y_s^2+z_s^2)+(x_r^2+y_r^2+z_r^2-T^2)]^2-4(x_s^2+y_s^2+z_s^2)(x_r^2+y_r^2+z_r^2).
\end{aligned}$$

Chapter 5

From constraints to inference

In this chapter we discuss how to reconstruct the geometry of an environment, using a set of controlled emissions and their related response acquisitions.

Choosing a standard audio equipment that limits ourselves to the audible range frequencies (20-20000 Hz), we have some drawbacks. First, the angular resolution of audible frequencies is low. Second, objects like walls are smooth with respect to the audible wavelengths, so sound is reflected and not diffused.

As said at the beginning of this thesis, we model the environment like a polyhedron (room). We assume that its faces are large and smooth with respect to wavelengths involved, in order to neglect diffusion and diffraction phenomena, and to apply acoustical optic theory. Under these hypotheses, we can exploit the image source method in order to estimate the behavior of the reflectors and then, we obtain a set of Times Of Arrival. Starting from these multiple measures and constraints, we explain how to turn them into a system of equations related to constraints. Next, we see in which way, in order to solve the system, we will deal with the minimization of the cost function coming from the system itself. The found minimum represents the estimate that the reconstruction process gives.

In the first section we investigate the theory behind the approach used to estimate a reflector in an environment.

In the second one, we analyze the cost function discussing its complexity and providing a description of a viable method to perform the minimization. In order to simplify this analysis, we introduce the Hough Transform, that can provide an help to this issue.

5.1 Estimation of a single reflector using multiple TOAs

As analyzed in Chapter 4, the set of tangent reflectors to a generic conic $\mathbf{C}$ is represented by the solutions of the equation

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

where $\mathbf{C}^*$ is the adjoint of the conic matrix $\mathbf{C}$, and $\mathbf{l}$ represents the vector parameter $[l_1, l_2, l_3, l_4]$ of a plane in $\mathbb{R}^3$.

As already mentioned, given a reflector and different positions of sources and receivers, we have one geometric constraint (ellipsoid) for each pair source-microphone. Consider a configuration source-receiver at the time instant m and another different configuration at the time instant k. Thus, the reflection points $P(m)$ and $P(k)$ will be displaced one from each other but they will be lie on the same reflector plane. Therefore, two ellipsoids $\mathbf{C}_m$ and $\mathbf{C}_k$ share one plane tangent to both of them (the reflector). An illustration is proposed in Figure 5.1. Consequently, the reflector can be estimated using the intersection of just two sets of tangent planes at different time intervals.

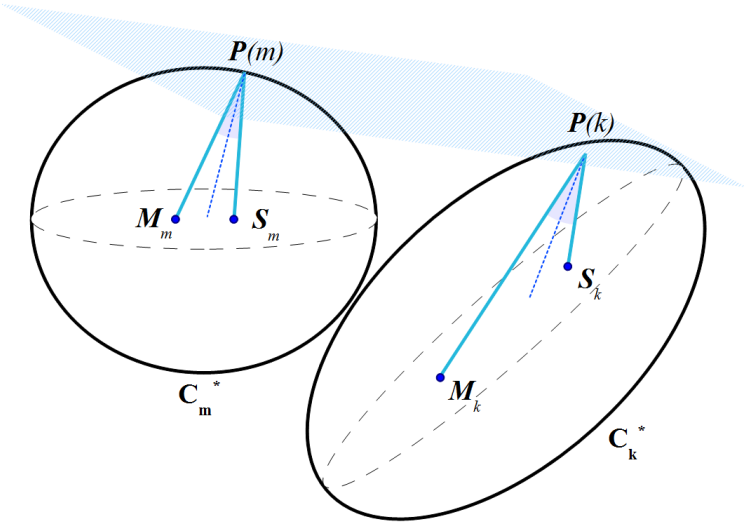


Figure 5.1: The tangent plane to both ellipsoids $\mathbf{C}_m^*$ and $\mathbf{C}_k^*$ is just the reflector.

In order to explain last concepts in mathematical terms, consider a single obstacle, one receiver placed in M, and S_k positions of loudspeakers, with

$k = 1, \ldots, K$. As just said the reflector can be inferred as the common tangent plane to the ellipsoids traced for S_k. More in detail, knowing foci S_k, M and the value of TOA τ_k^r estimated by the inspection of $h_k(t)$, we are able to trace the ellipsoids $\mathbf{C}_k$, so we can write

$$\begin{cases} \mathbf{l}^T \mathbf{C}_1^* \mathbf{l} = 0 \\ \mathbf{l}^T \mathbf{C}_2^* \mathbf{l} = 0 \\ \quad \ldots \\ \mathbf{l}^T \mathbf{C}_K^* \mathbf{l} = 0 \end{cases} \tag{5.1}$$

Since we have four unknowns l_1, l_2, l_3, l_4, we need at least four equations in the system.

Now we have to do two considerations. First, since matrices $\mathbf{C}_1, \ldots \mathbf{C}_K$ are full-rank (except for degenerate conics) we cannot find the solution of the system as the intersection of the nullspaces of $\mathbf{C}_k$, because each nullspace is empty. Second, and more important thing is that if $\mathbf{l}$ is a solution of the system, then also $k\mathbf{l}$, with $k \in \mathbb{R}$, is so, so we could look for the solution in a suitable subspace of $\mathbb{P}^3$ where an unique solution is present. The latter consideration will be useful in the next section, about minimization of the cost function.

5.2 The minimization problem

As explained in the last section, our aim is to find the solution of the system (5.1). This issue is not trivial dealing with nonlinear systems. In fact, measures of TOAs are affected by estimation errors, and this can lead the system to have no exact solution. Hence, we will use an approach based on nonlinear Least Squares, trying to minimize the cost function

$$J(\mathbf{l}) = \sum_{k=1}^{K} (\mathbf{l}^T \mathbf{C}_k^* \mathbf{l})^2, \tag{5.2}$$

and looking for the parameter vector $\hat{\mathbf{l}}$ that minimizes J

$$\hat{\mathbf{l}} = \arg\min_{\mathbf{l}} J(\mathbf{l}). \tag{5.3}$$

Now, we make some observations. Matrices $\mathbf{C}_k^*$ are neither negative nor positive definite: in fact an ellipsoid in $\mathbb{P}^3$ is represented by a matrix with three

positive eigenvalues, while the fourth one is negative. Hence, the function in (5.3) is not granted to have a global minimum.

Moreover, as noticed in the last section, because of the scalability of the vectors in $\mathbb{P}^3$, the system has infinite solutions, so we need to impose a normalization of the constraints. This fact turns into considering a convenient subspace of $\mathbb{R}^4$ in which the minimization problem is granted to have a unique solution.

In order to limit the search space, we look for unit norm solutions $||\mathbf{l}|| = \mathbf{l}^T\mathbf{l} = 1$. Therefore, we now search the solution of the constrained minimization problem

$$\hat{\mathbf{l}} = \arg\min_{\mathbf{l}} J(\mathbf{l}) \quad \text{subject to} \quad \mathbf{l}^T\mathbf{l} = 1.$$

The constraint we have just introduced, corresponds to looking for solutions the surface of an hypersphere of unitary radius. In fact, the parameter vector $\mathbf{l} = [l_1, l_2, l_3, l_4]$ need to be described in a four-dimensional space, and so we use hyperspherical coordinates.

A generic point on an hypersphere is defined as

$$\begin{cases} kl_1 = \rho\sin(\theta)\sin(\phi)\sin(\alpha) & = \sin(\theta)\sin(\phi)\sin(\alpha) \\ kl_2 = \rho\sin(\theta)\sin(\phi)\cos(\alpha) & = \sin(\theta)\sin(\phi)\cos(\alpha) \\ kl_3 = \rho\sin(\theta)\cos(\phi) & = \sin(\theta)\cos(\phi) \\ kl_4 = \rho\cos(\theta) & = \cos(\theta) \end{cases}$$

where, in the second equation, we have considered our constraint, i.e. the unitary radius.

In this hypothesis, the minimization problem turns into finding the new solution

$$\hat{\mathbf{r}} = \arg\min_{\mathbf{r}} J(\mathbf{r}),$$

with

$$\mathbf{r} = [\theta, \phi, \alpha]^T, \quad 0 < \theta < \pi, \quad 0 < \phi < 2\pi, \quad 0 < \alpha < 2\pi.$$

However, the cost function does not result a trivial object to study, just because of the presence of several local minima instead of a global minimum. Thus, one approach to simplify this analysis, is to provide to the non-linear Least Squares algorithm an initial guess of the reflector position. Doing so, the minimization task is more usable. The technique we use for this aim derives from an algorithm known in the community of computer vision as the Hough Transform.

5.3 Initialization using the Hough Transform

In this section we discuss how we can drive the minimization process of the cost function.

The Hough transform is a technique that allows detection of global configurations present in an image. It is used in image analysis, computer vision and digital image processing. The Hough transform as it is universally used today was invented by R. Duda and P. Hart in 1972 ([34]), who called it a *generalized Hough transform* after the related 1962 patent of P. Hough ([35]).

Since it requires that the desired features are specified in some parametric form, the *classical Hough transform* was concerned with the identification of lines, and later extended to identifying positions of arbitrary shapes, most commonly circles or ellipses. Conversely, the *generalized Hough transform* can be employed in applications where a simple analytic description of a features is not possible.

The purpose of the technique is to find instances of objects within a certain class of shapes by a voting procedure. This voting procedure is carried out in a *parameter space*, from which object candidates are obtained as local maxima in a accumulator space, and it is performed by a *validation of hypotheses*. The accumulator space is explicitly constructed by the algorithm for computing the Hough transform.

5.3.1 The parameter space

Consider one microphone and one source. For simplicity of computation consider the microphone placed in the origin of reference system (however, calculus can be easily generalized to different hypotheses). Consider the presence of one reflector. We define these three objects by using spherical coordinates. In particular the source is represented by ρ_k, θ_k, ϕ_k, with k the lag referred to the k-th position of the loudspeaker. The position of a generic reflector is identified by the coordinates of the normal to the plane ρ, θ, ϕ.

According to considerations in Chapter 2 and Chapter 4, we can compute the TOA τ_k as the travel time of the reflected ray, limited to first-order reflections, from the source to the microphone.

Graphically the Time of Arrival and the reflection point can be obtained by using the image source method (Chapter 2). A representation is given in Figure 5.2. Depending on our hypothesis, it is convenient to take the

mirrored image of the microphone respect to the plane. So, the time of arrival is inferred by computing the distance between the loudspeaker and the image source just found. Thus, we can easily write

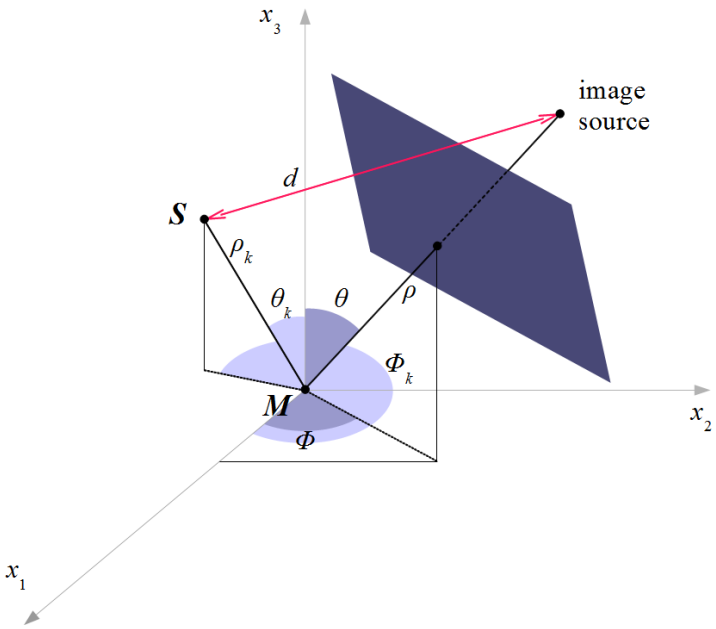


Figure 5.2: TOAs are computed through the distance d between the source and the image source respect to the microphone (or viceversa) by using spherical coordinates: ρ, θ, ϕ for the reflector and ρ_k, θ_k, ϕ_k for the k-th position of the source.

$$\begin{aligned}\tau_k(\rho,\theta,\phi) =& \frac{1}{c}\big[(2\rho\sin\theta\cos\phi - \rho_k\sin\theta_k\cos\phi_k)^2 \\ &+ (2\rho\sin\theta\sin\phi - \rho_k\sin\theta_k\sin\phi_k)^2 + (2\rho\cos\theta - \rho_k\cos\theta_k)^2\big]^{\frac{1}{2}},\end{aligned} \tag{5.4}$$

where c is the propagation speed. After some computation equation (5.4) becomes

$$\begin{aligned}\tau_k &= \frac{1}{c}2\rho\sqrt{4\rho^2 + \rho_k^2 - 2\rho\rho_k\cos(\phi-\phi_k)\sin\theta\sin\theta_k} \\ &= \frac{1}{c}\big[2\rho - \frac{\rho_k}{2}\cos(\phi-\phi_k)\sin\theta\sin\theta_k\big],\end{aligned}$$

where in the second equality we used a Taylor approximation of the first order and we neglected the term $\frac{\rho_k^2}{4\rho}$ since his contribution is irrelevant when the obstacle is distant ($\rho >> \rho_k$).

This equation relates the path length of a first-order reflected signal, i.e the value of the corresponding TOA, to the positions of involved source and reflector, having assumed the microphone placed in the origin. Anyway, this last assumption has been introduced just in order to make less complex the treatment. In fact, given the position of the microphone, it is sufficient to shift the reference system in one which has the receiver in the origin. This is realized mathematically subtracting to each term related to the component x, y, z in (5.4), the corresponding term identifying the position of the microphone. So, after calculus that we omit, the previous equation becomes

$$\begin{aligned}\tau_k(\rho,\theta,\phi) = \frac{1}{c}\big[&(2\rho\sin\theta\cos\phi - \rho_k\sin\theta_k\cos\phi_k - \rho_m\sin\theta_m\cos\phi_m)^2 \\ &+ (2\rho\sin\theta\sin\phi - \rho_k\sin\theta_k\sin\phi_k - \rho_m\sin\theta_m\sin\phi_m)^2 \\ &+ (2\rho\cos\theta - \rho_k\cos\theta_k - \rho_m\cos\theta_m)^2\big]^{\frac{1}{2}},\end{aligned}$$

where ρ_m, θ_m, ϕ_m are the spherical coordinates of the microphone location.

In conclusion, given the position coordinates of the loudspeaker and the receiver, a specific value of TOA generated by a reflector, is computed through a triplet of parameters ρ, θ, ϕ identifying just the reflector position. This is the reason for which working with the Hough transform we talk about a parameter space, constituted by possible values of these radial and angular coordinates.

5.3.2 The voting procedure

In the last paragraph we have seen how the values of TOAs are related to a parameter space of ρ, θ, ϕ, where a combination of these variables identify a position of the reflector. The Hough transform acts on localization task, through an hypotheses validation. We now analyze it.

Defined a value of TOA we want to look for in the scene, for each point of the image we compute the three parameters of all planes (reflectors) that can produce that specific TOA and we upgrade the specific cells of a 3D space corresponding to the various reflectors. We thus obtain an accumulation function defined in the parameter space. Therefore, local maxima of this function, i.e. points in the parameter space that count the larger number of

votes, represent the objects most likely to be present in the image, as if it were an hypothesis supported by many experimental confirmations (in fact in many cases the approach is called *hypotheses validation*).

We propose a practical example. In our situation, the set of TOA from which start, is found looking for peaks of the cross-correlation function. Considering an acquired value of TOA $\hat{\tau}$, we can compare it with the value of TOA τ_k computed putting in (5.4) ρ, θ, ϕ from a range of possible values, so guessing a reflector. Then, we have to check if $|\hat{\tau} - \tau_k| < \epsilon$, with ϵ a small acceptance threshold. If this happens, it means that the two TOAs are nearly equal, and likely the reflector we have guessed is the actual reflector from which $\hat{\tau}$ has been computed. The task is repeated for all the combinations of values in the parameter space.

Dealing with a single reflector in the environment, the 3D map will show a cell more voted respect to the others.

We propose a short example which can better describe the accumulator space. A single reflector (plane) in the scene is identified by the vector parameter

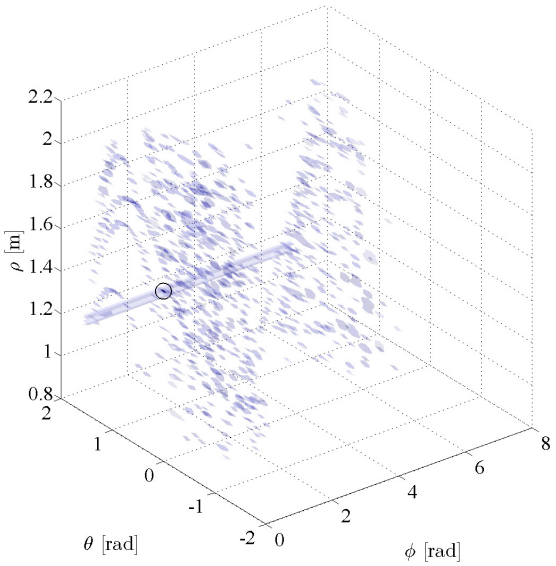
$$=[1, 0, 0, -1.6],$$

thus its spherical coordinates are

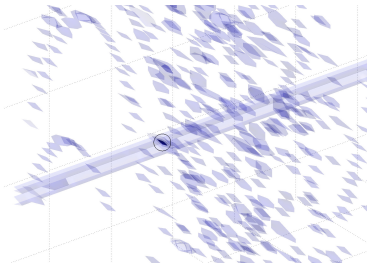
$$\rho = 1.6, \qquad \theta = 0, \qquad \phi = 0.$$

In Figure 5.3 the resulting Hough map is shown. The estimate provided by the Hough Transform shows in the figure a very small region, represented in dark blue, most likely referred to one cell or a group of cells more voted respect to the others. The position of the actual plane is marked with a circle. In fact, the larger number of votes is concentrated in this region. It is so small because we have chosen to use a large number of values for ρ, θ, ϕ, thus it is more difficult to distinguish the point in so wide ranges. Moreover, the presence of few dark points, even if less not so visible, nearby the circled one, is due to the fact that we chose to represent in the graph not only the maximum of the map, but also points whose values are close to it.

Summarizing, by using this approach, we can obtain an initial guess of reflector position, such that the minimization of the cost function could be driven and the estimate less difficult to perform. In fact, in the situation in which the algorithm extracts from the map not a single point, but a cluster



(a) *Hough map.*



(b) *Zoom on the most voted cell.*

Figure 5.3: Hough map: cells more voted than others are represented in darkest blue. Light shades are referred to all points under the maximum value: if the number of votes increases also the color intensifies.

of points, i.e. multiple cells voted more or less the same number of times, one close to each other and concentrated nearby of a local maximum, we have to adopt some strategies to turn the cluster into a single reflector. Existing techniques are more that one: it is possible to make an average of planes in a cluster (centroid), or to choose the plane that minimizes the cost function, etc.

We analyze in the next section, how the minimization works and why a starting point is useful for the process.

5.4 The non-linear Least Squares method

As we said in the end of the last section, the Hough Transform can provide by itself a right estimate for the reconstruction task, but it can also be useful to drive the minimization of the cost function. Now we explain in which way this happens.

The approach we follow to solve the system (5.1) is the non-linear Least Squares method. The goal of this method is to find $x \in \mathbb{R}^n$ that minimizes

$$\left\|r(x)^2\right\| = \sum_{i=1}^{m} r_i(x)^2, \tag{5.5}$$

where $r : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$ and r_i is a nonlinear function of the n-vector of variables x. Equation (5.1) can be seen as

$$\begin{aligned} r_1(x_1, x_2, \ldots, x_n) &= 0 \\ r_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ r_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

In general, $g(x) = \left\|r(x)\right\|^2$ may have multiple local minima and finding the global minimum is usually very hard, as widely discussed in previous sections. The nonlinear minimization problems require an iterative approach. There are several methods to solve this kinds of problems. We shortly describe one of these, the *Gauss-Newton* method.

It is based on implementing first-order derivatives of the components of the vector function and on a linear approximation to the components of r in

the neighbourhood of x. In practice, we start with an initial guess $x^{(0)}$. At the iteration k, we linearize $r_i(x)$ around current guess $x^{(k)}$ as we show:

$$r_i(x) \approx r_i(x) + \nabla r_i(x^{(k)})^T (x - x^{(k)}).$$

Now, we act minimizing

$$\sum_{i=1}^{m} \Big(r_i(x) \approx r_i(x) + \nabla r_i(x^{(k)})^T (x - x^{(k)}) \Big)^2,$$

i.e. the sum of the squares of the linearized residuals and where (∇r_i) is the Jacobian: $(\nabla r_i)_{ij} = \frac{\partial r_i}{\partial x_j}$. Now, the goal is to find the new guess $x^{(k+1)}$ from $x^{(k)}$ solving a linear least-squares problem. Hence, we minimize

$$\left\| A^{(k)} x - b^{(k)} \right\|^2,$$

with

$$A^{(k)} = \begin{bmatrix} \nabla r_1(x^{(k)})^T \\ \nabla r_2(x^{(k)})^T \\ \vdots \\ \nabla r_m(x^{(k)})^T \end{bmatrix}, \qquad b^{(k)} = \begin{bmatrix} \nabla r_1(x^{(k)})^T x^{(k)} - r_1(x^{(k)}) \\ \nabla r_2(x^{(k)})^T x^{(k)} - r_2(x^{(k)}) \\ \vdots \\ \nabla r_m(x^{(k)})^T x^{(k) - r_m(x^{(k)})} \end{bmatrix}.$$

The new guess $x^{(k+1)}$ is found by using the pseudo-inverse of matrix $A^{(k)}$. Hence

$$x^{(k+1)} = \left(A^{(k)T} A^{(k)} \right)^{-1} A^{(k)T} b^{(k)}.$$

The steps are repeated until convergence, although it is not guaranteed: the behavior of the iteration algorithm depends very strongly on the initial estimate $x^{(0)}$.

In conclusion, what we do through the Hough Transform, is to estimate a point which can be used as initial guess of the nonlinear minimization process, that here we call $x^{(0)}$. Doing so, we allow to drive the minimization process make more usable the cost function in (5.2). This kind of strategy, moreover, is such that the final result is represented by a global minimum, instead of multiple local maxima.

Chapter 6

Extension to multiple reflectors

As we saw in the Chapter 5, the Hough Transform can provide an initial guess for the minimization process, making it easier to compute. Acting so, we can obtain a global minimum instead of several local minima.

In this chapter we discuss how the Hough transform, can be used also in the situation of multiple reflectors and it can provide an estimate of all obstacles in the scene, i.e. the geometry of the environment. In the first section we provide the data model representing a situation with multiple involved reflectors. In the second one, we describe in detail the extension of the Hough Transform to estimate multiple planes, discussing in detail some important aspects concerning with the correct working of the method. Finally, it is proposed a practical example of an Hough map in the case of multiple reflectors, to summarize all considerations before.

6.1 Data model

In the situation of multiple reflectors, we can not apply the techniques we saw before, as the minimization through the common tangent plane to ellipsoids. In fact, with multiple obstacles, the microphone receives different replicas of the signal due just to multiple reflections. Now we define the model and how we study it.

Consider some reflectors in an environment, each one of them is identified by the lag $g = 1, \dots, G$. Our aim is to estimate planes $\hat{\mathbf{p}}_g$ on which each reflector lies. In the general case of mutually visible reflectors, the received signal will exhibits not only first-order reflections but also high-order

reflections.

If the emitted signal is, as before, a time-continuous sequence, the acquired signal at the time instant k is

$$x(t,k) = \alpha_d(k)s(t-\tau_d(k)) + \sum_{g=1}^{G} \alpha_g(k)s(t-\tau_g(k)) + \sum_{h=1}^{H} \alpha_h(k)s(t-\tau_h(k)) + \nu(t),$$

where lag d, g, h are referred to direct signal, first-order and high-order reflections, respectively.

The cross-correlation between the received signal and the reflected one at the time k is

$$h(t) = \alpha_d(k)\delta(t-\tau_d(k)) + \sum_{g=1}^{G} \alpha_g(k)\delta(t-\tau_g(k)) + \sum_{h=1}^{H} \alpha_h(k)\delta(t-\tau_h(k)) + r_\nu(t).$$

In these hypothesis, TOAs referred to reflectors $g = 1, \ldots, G$ (we consider only first-order reflections) are mixed together. In fact, the impulse response just computed, exhibits multiple local maxima, each one of them represents the value τ_{kj}, but they are not organized in a labeled dataset, so we cannot know which reflector the element τ_{kj} is assigned to.

Therefore our aim is to apply a strategy that transforms the dataset τ_{kj} in a labeled dataset $\hat{\tau}_g(k)$, $g = 1, \ldots, G$. After having found TOAs belonging to the same reflector, we would be able to apply the common tangent algorithm.

The approach we exploit is to describe the problem of inferring all reflectors in the environment as the superposition of smaller problems, in which we try to find a reflector at once.

6.2 Labeling TOAs by the Hough transform

We found in the Chapter 5, that a simple peak-picking of the cross-correlation function returns the values of TOAs: if we are in the situation of a single reflector, the impulse response will exhibit one peak for the direct signal and one for the first-order reflection, while if we are dealing with multiple reflectors the cross-correlation function will contain several significant peaks together with the peak of direct signal. In this last case, we have not the knowledge about which obstacle each TOA of the reflected signal is generated from. So, as mentioned in the previous section, the main problem, is to assign

TOAs of the dataset acquired, to the correct reflector which they are referred to.

This issue, in a situation of multiple reflectors, turns into establishing if τ_{kj} is related to τ_{mi} with $k \neq m$ and $i \neq j$, based on the fact that they are both related to the same reflector. Moreover, we cannot use amplitude for peak tracking because amplitude of peaks can be time-dependent.

In practice, we can exploit the Hough transform also for this purpose: the approach we analyzed previously in order to infer a single obstacle based on the validation of hypotheses theory, can be easily extended to the situation of multiple obstacles. Recall the Figure 5.2 and rewrite equation (5.4):

$$\begin{aligned}\tau_g(k)(\rho,\theta,\phi) =& \frac{1}{c}\big[(2\rho_g \sin\theta_g \cos\phi_g - \rho_k \sin\theta_k \cos\phi_k)^2 \\ &+ (2\rho_g \sin\theta_g \sin\phi_g - \rho_k \sin\theta_k \sin\phi_k)^2 \\ &+ (2\rho_g \cos\theta_g - \rho_k \cos\theta_k)^2\big]^{\frac{1}{2}},\end{aligned} \tag{6.1}$$

which describes the TOA computed for a reflector defined by ρ_g, θ_g, ϕ_g, a loudspeaker represented by ρ_k, θ_k, ϕ_k, a microphone in the origin of the reference system, and where we have computed the image source with respect to the microphone.

Hence, the task of TOAs assignment for multiple reflectors, turns into finding the set of parameters $\{\rho_g, \theta_g, \phi_g\}$, with $g = 1, \ldots, G$, that best fits the dataset τ_{kj}, $j = 1, \ldots, J$, $k = 1, \ldots, K$. Once the points are classified, they are assigned to the reflector g to estimate the function $\hat{\tau}_g(k)$, $g = 1, \ldots, G$. After the labeling, it is possible to apply the common tangent algorithm, since now we can use TOAs grouped for each reflector.

As already explained in Chapter 5, the task is realized through a sort of map. This is represented by a n-dimensional grid of cells where n is the number of parameters. In our case the parameter space is reduced to a three-dimensional space, the map $\boldsymbol{m}(\rho_g, \theta_g, \phi_g)$. Each intersection point in the map represents a combination of three values of ρ, θ, ϕ, so the map contains all possible reflectors we consider in our estimate process. The local maxima are selected as the most likely reflectors $\{\hat{\rho}_g, \hat{\theta}_g, \hat{\phi}_g\}$. Therefore, given the reflectors $\{\hat{\rho}_g, \hat{\theta}_g, \hat{\phi}_g\}$, $g = 1, \ldots, G$, their TOAs must lie on $\hat{\tau}_g(k)$ according to the (6.1). Now, the assignment of the TOA τ_{kj} to the reflector g occurs

in this way:

$$\tau_g(k) = \begin{cases} \arg\min_j |\tau_{kj} - \hat{\tau}_g(k)| & \text{if } \min|\tau_{kj} - \hat{\tau}_g(k)| < \epsilon \\ \hat{\tau}_g(k) & \text{if } \min|\tau_{kj} - \hat{\tau}_g(k)| > \epsilon \end{cases} \tag{6.2}$$

where ϵ is an acceptance threshold. We notice that not all τ_{kj} are necessarily assigned to a reflector.

Now we move to discuss how the method works in detail, providing in the end, an example of Hough map with multiple involved reflectors.

6.2.1 The Hough map: involved parameters

We would focus on some important aspects of the approach by the Hough Transform. We are interested to explain better how the map is structured and what are the parameters that the method takes into account.

The approach by Hough transform, is structured in three main steps, each of which involves specific parameters. An illustration is shown in Table 6.1.

Table 6.1: The Hough transform: the main steps and involved variables.

	Goal	*Parameters*
Step 1	Create the empty accumulator space setting a size for each of three parameter ranges.	Min/max values, stepsizes
Step 2	Compute TOAs, compare them with acquired ones and start with voting procedure.	Acceptance threshold ϵ
Step 3	Select from the map the most voted cells.	Selection threshold

Step 1

Each parameter ρ, θ, ϕ, constituting one dimension of the the Hough map (the accumulator space), assumes values in a range identified by a minimum/maximum value and a stepsize. How do we have to choose them? The

end points of ranges bound a region of space in which we think reflectors will be estimated. The stepsize, instead, is responsible for the number of values that the algorithm takes in input, that is how much plane positions are guessed and how much the map is crowded with values. Hence, both these parameters must be chosen carefully: whether actual reflectors are not represented by any combination of values from ranges, the algorithm localizes nothing. On the other hand we should avoid waste of computation, trying not to make wider ranges than necessary if, for instance, we have an idea about geometry of the environment. In fact, in real cases this is not applicable. Also the value of stepsize, on the other side, can affect the faithfulness of the estimate, mainly in real experiments or complex positions of reflectors. In fact, in order to obtain a correct estimate, we must be sure that actual planes in the environment will be taken into account in the Hough map. If this does not happen, the algorithm searches the closest plane to the missing one, whose values of TOAs are more similar to those of the actual reflector. Thus, in the luckiest case, the estimate is something closed to real geometry; in the worst case (too few values) the algorithm is unable to locate anything. This ambiguity can be reduced filling the ranges with a dense distribution of values. Obviously this implies a trade-off between faithfulness and computation complexity.

Step 2

The voting procedure takes place on the base of a comparison between acquired TOAs and those computed by the algorithm. This task is realized checking if the module of the difference between the two values of TOAs to compare, is under an acceptance threshold ϵ. If we take the value of ϵ too low, it most likely means that measures are really faithful and precise, because we are expecting that TOAs acquired are almost identical to those computed. Viceversa, if ϵ is chosen too high, we can risk to include in the estimate also reflectors that are not present in the scene. Thus, a good acceptance threshold has to be a trade-off, depending mainly on the situation we meet.

Step 3

The estimate that the Hough Transform provides, is obtained selecting points in the map. This selection is performed by extracting cells that count a num-

ber of votes above a threshold. How do we have to set this parameter? This is not a trivial choice. Given some sources and some receivers, the acquired TOAs for each reflector are in number as the product between the number of sources and that of microphones, i.e the number of acquisitions. Therefore, the more voted cell in the Hough map can be counted at most as the number of acquisitions. However, this is not frequent, since it depends strongly on several conditions as noise that affects measurements or the complexity of plane positions that does not help to know whether the algorithm will take in account or not of those specific planes. Hence, in general this threshold is not chosen to extract only the cells counting the maximum value, risking in this way to take nothing. The selection threshold is usually set at lower values than the number of acquisitions, for instance the 50% or 40% of it. Lower threshold, in general, do not get worse the estimate, but rather allow to localize clusters of points, i.e. groups of cells, one close to each other, voted approximately the same number of times. On the other hand, decreasing the selection threshold too much, we risk to obtain an estimate that includes other reflectors in regions well away from those of the actual geometry. Nevertheless, there exist some techniques that could be able to discard points irrelevant to inference, trying to analyze the concentrations of clusters (the richer ones will be the actual ones).

6.2.2 The Hough map: a practical example

In this paragraph we want to provide an example of Hough map, illustrating the effects of the voting procedure in the situation of multiple obstacles.

In the case of a single reflector in the environment, the accumulator space will most likely exhibit a cell or a cluster of cells more voted with respect to the others. Dealing instead with multiple objects in the environment, the accumulation function will contain multiple clusters of points or multiple points, identifying different regions of space, close to the positions of the objects in the scene. We provide an example, also with the aim to show how the estimate is mapped from the space parameter $\{\rho, \theta, \phi\}$ in which reflectors are represented by spherical coordinates to that $\{x, y, z\}$ in which Euclidian coordinates are identify planes.

Consider three reflectors in the environment, whose the equations are described by vectors parameters

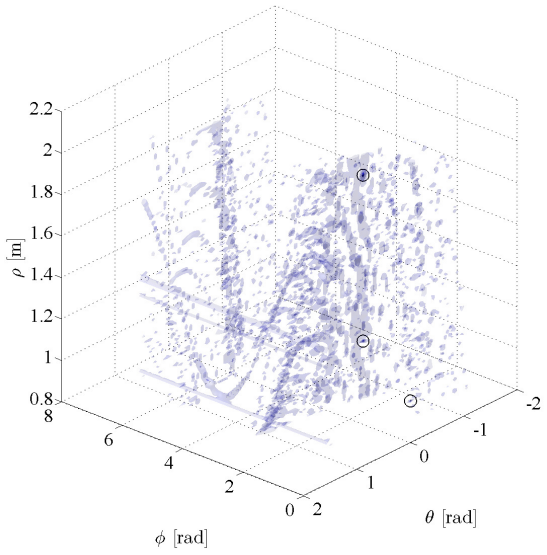
$$\mathbf{l}_1 = [1, 0, 0, -1], \qquad \mathbf{l}_2 = [0, 1, 0, -1.2], \qquad \mathbf{l}_3 = [0, 1, 0, -2].$$

Applying the algorithm, the resulting Hough map and the estimate extracted from it, are those reported in Figure 6.1 (which includes some zooms on maxima). The Hough map shows three main darkest regions, representing the cells or clusters of cells more voted with respect to the others. The positions of the actual planes are marked with a circle.

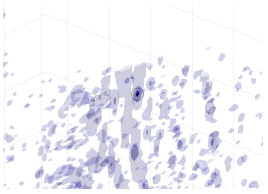
The obtained results can be used to initialize the minimization task. We have to adopt the same kind of strategy mentioned in Chapter 5 (centroid, plane that minimizes the cost function, etc) to obtain a single minimum as starting point.

Sometimes, just the algorithm referred to the Hough transform can include a part of code able to check if the planes can be incorporated in a single plane or not, in order to reduce the density of the estimate, making simpler the computation. There exist several techniques to do it. The strategy we adopt is to start taking a reflector, to compute its distance from the others, and if it is sufficiently small, to include them in the first considered reflector. The method is repeated for each localized plane. This could make the estimate more or less identical to the real geometry, but still good. However, this is not a needful task.

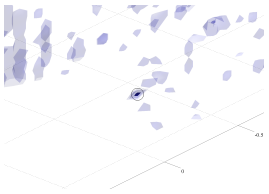
In conclusion, even though the Hough transform was not able to provide the exact position of the objects, it could still give an acceptable estimate, finding regions of concentration in space in which planes most likely lie. Therefore, the Hough transform even without the minimization task, gives an initial estimate of the environment geometry.



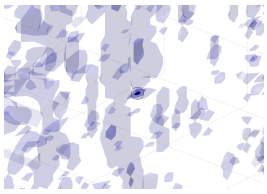
(a) *Hough map.*



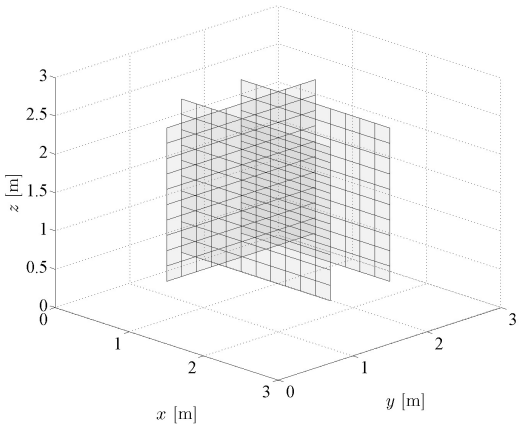
(b) 1° *maximum.*



(c) 2° *maximum.*



(d) 3° *maximum.*



(e) *Estimated reflectors.*

Figure 6.1: We see which is the relation between the voting procedure in the space parameter $\{\rho, \theta, \phi\}$ and the localization of reflectors in the space $\{x, y, z\}$.

Chapter 7

Validation

In this chapter we show the results of our simulations and real experiments. We also analyze different aspects of the approach not considered in the theoretical analysis: for instance we will see in which measure noise can influence a correct estimate, and as a consequence, how this fact makes necessary to apply some changes to the algorithm.

The presented simulations aim to analyze the robustness of the algorithm, especially how it responds to changes of conditions or variables involved, like parameters in the Hough transform, number of sources or microphones, addition of noise, etc.

About the real experiments, instead, we carried out two kinds of tests. The first one takes place in an anechoic room: we use some not muffling panels in order to create reflected rays and placing them in number and positions depending on the complexity of the experiment and the effects we want to simulate. On the contrary, the second kind of experiments, is performed in a common room, in order to take in account every sort of involved effect and to evaluate the performance of our algorithm in a real and more common situation.

One difference between simulations and real experiments, is the way through which we obtain TOAs with which to compare those computed in the Hough transform. While in simulations, Times Of Arrival are obtained involving mathematics, so not including noise, since they are computed exactly, dealing with real experiments they are acquired by picking peaks from the impulse responses. Therefore, we can produce more realistic results even in simulations, by adding, to calculated TOAs, a random value representing noise.

Before we discuss performed tests, we would give again a short survey about the steps constituting the reconstruction process, providing for each one of them a graphical example which shows what, the implementation of of these tasks, really means.

Recalling the Figure 4.1 describing the idea behind of the acoustic scene

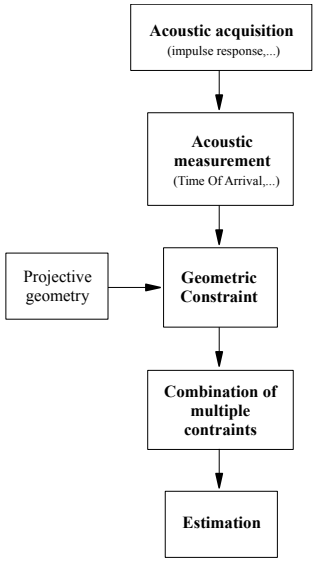


Figure 7.1: General scheme of a geometric analysis approach.

reconstruction, go back over each step, providing a more detailed explanation.

Any acoustic reconstruction process starts with the acquisitions of a signal. Thus it needs microphones and sources, through which the emitted signal can be recorded. Given the position of one source, the acquired signal reaches each receiver after a time delay, both for the direct paths and the reflected ones. Thus, given a source, performing the cross-correlation between the recorded signal for each acquisition (for each reached microphone) and the emitted signal (white noise), we obtain the impulse response containing all time delays involving the specific pair source-receiver (direct paths, first-order reflections, second-order reflections,etc). Therefore, neglecting the time delay relative to the direct signal, which has the shortest propagation time, all TOAs referred to reflected paths can be extracted by selecting local maxima in the impulse response. In order to discriminate the first-order reflections from the high-order ones, and understand from which reflectors they have been generated, we have to exploit the joint analysis of the impulse responses. In Figure 7.2 an example of impulse response is proposed.

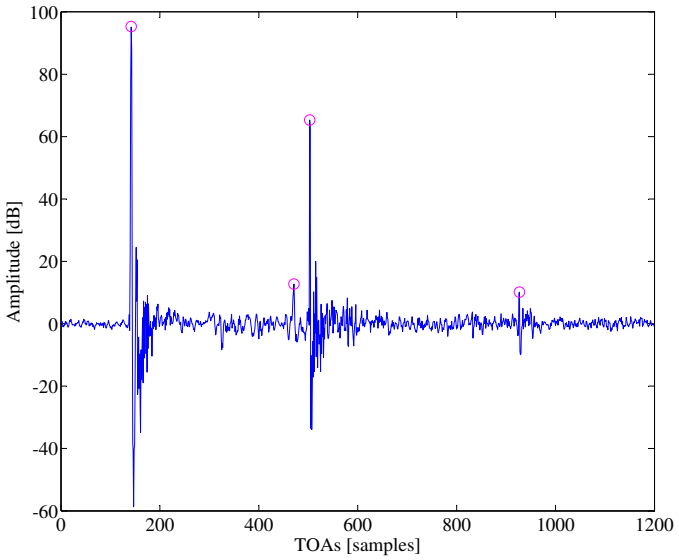


Figure 7.2: Cross-correlation between the emitted signal (white noise) and the acquired one. Peaks are circled.

Through the extraction of TOAs from the impulse responses, we obtain a unlabeled dataset of Times Of Arrivals. A representation is given in Figure 7.3. We can observe that some TOAs appears for some acquisitions and disappear for others. Then, our aim is to label this dataset, so assign each Time Of Arrival to the reflector which generated it. This task is accomplished using local maxima in the Hough map.

Subsequently, we involve the projective geometry in order to study the constraints which derive from the first-order paths, whose we discussed in several parts of this thesis. Particulary, given the positions of a source and a receiver, a specific value of TOA constrains the reflection points to lie on an ellipsoid just with the source and the microphone as foci. An illustration is given in Figure 7.4: given a source, a receiver, and a value of TOA, together with an obstacle, the reflected path has to honor the Snell's law of reflection. The blue symbols represent foci (source and receiver), while the magenta one, is referred to the image source of the source respect to the reflector.

The last step of the reconstruction process is performed combining multiple constraints derived by TOAs. Using the labeled set of TOAs, referring to each reflector in the scene, we combine the equations describing all lines

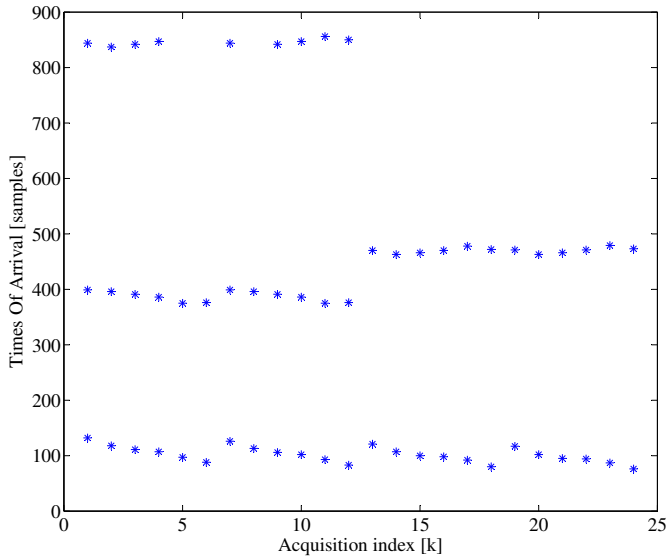


Figure 7.3: Times Of Arrival extracted by the cross-correlation function at each acquisition before labeling.

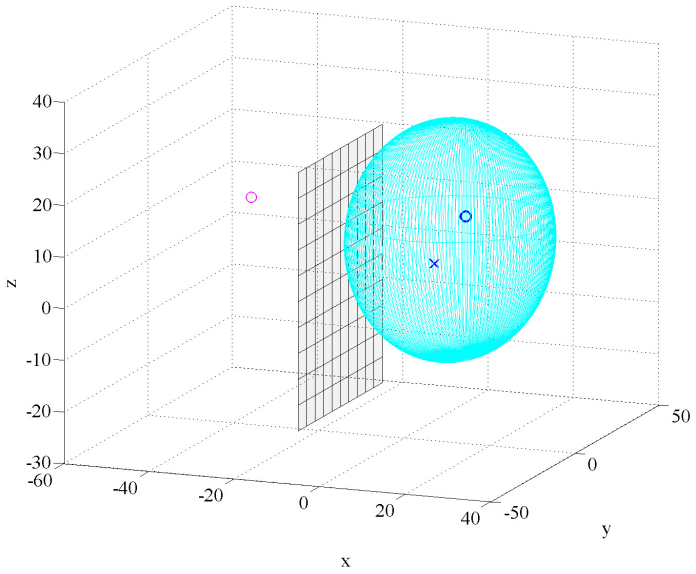


Figure 7.4: A geometric constraint derived from source, receiver and reflector location.

tangent to each ellipsoid obtained by any combination source-microphone, in a system of equations, one for reflector, in order to find the vectors parameter $\mathbf{l} = [l_1, l_2, l_3, l_4]^T$ identifying obstacles in the scene. In this way we are minimizing the resulting cost function. The used approach is the non-linear Least Squares by iterative method, thus we feed an initial guess of minimum coming from the estimate provided by the Hough Transform. The resulting global minimum/minima (single/multiple reflectors) allows to infer geometry. An example of cost function is provided in Figure 7.5. It is represented by isosurfaces, since we aim to show how the cost function changes in the space parameter identified by θ, ϕ, α, needful to represent any vector parameter $\mathbf{l} = [l_1, l_2, l_3, l_4]^T$: the hyperspherical coordinates are necessary to represent a four-dimensional space. The blue circle is referred to the global minimum estimated by the minimization task.

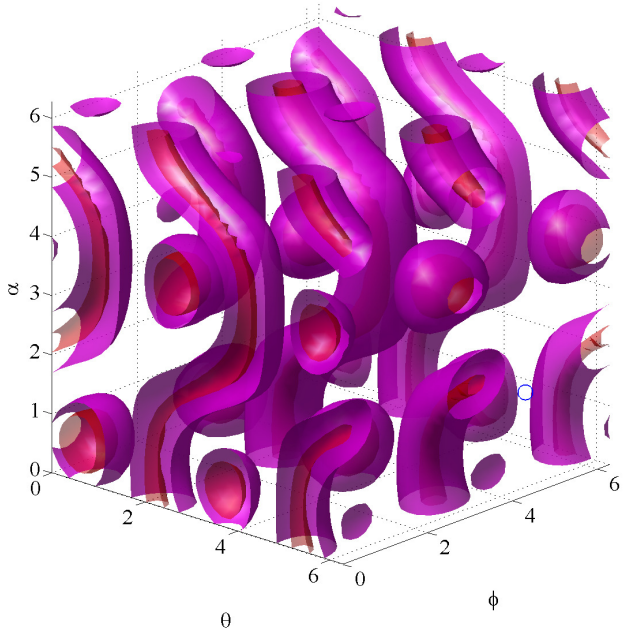


Figure 7.5: A cost function for any combination of parameters θ, ϕ, α. The global minimum is identified by the blue circle.

7.1 Simulations

In this section we analyze several simulations showing their results, with the goal to validate principles provided in Chapters 5 and 6.

We subdivide tests by changing parameters or conditions. We are going to analyze the robustness of the algorithm, particulary regarding to the behavior of the Hough Transform, because it alone can provide all needful information from which start inferring the geometry through the minimization task. Hence, we propose some simulations that describe how the Hough Transform behaves if we change the value of some involved parameters.

Moreover, in almost all graphs proposed in the following, widths, heights and in many cases also positions of planes may be different one to each other, just for reason of layout, and in order to try an easier and immediate idea of localization. In fact, it is maybe obvious but important to specify, that reflectors we consider are planes with infinite extension, and not faces of 3D objects, as instead design would suggest.

We start with analying three main parameters at the base of the Hough algorithm: acceptance threshold, size of the space parameter and selection threshold.

7.1.1 Varying the acceptance threshold

As widely explained in the previous chapters, we saw that the Hough algorithm is characterized by different parameters. One of this is the threshold ϵ, that defines an acceptance range in samples within which two values of TOAs are considered equal. As already said, this task allows the voting procedure, in which the counter of a cell is incremented if the TOA $\hat{\tau}$ computed referring to the reflector identified by that cell, is such that $|\tau - \hat{\tau}| < \epsilon$, where τ is the value of the acquired Time Of Arrival in samples.

We perform a simulation considering a single plane in the room. For simplicity consider a vertical plane. We consider also five sources and three microphones. We test different values of increasing threshold ϵ, keeping constant the other involved parameters whose values are shown in Table 7.1 and Table 7.2.

We choose $\epsilon = 1, 3, 8$ samples. The results are shown in Figure 7.6. For the smallest acceptance threshold the algorithm localizes one plane, colored in gray, perfectly overlapped to our actual plane, colored in blue. With $\epsilon = 3$

Table 7.1: Space parameter $\{\phi, \theta, \rho\}$.

Parameter	Min value	Max value	Stepsize
ρ[m]	0.7	1.2	0.01
$\theta[^\circ]$	-90	90	10
$\phi[^\circ]$	0	355.5	4.5

Table 7.2: Other parameters: sources, receivers, thresholds.

Parameter	Value
Sources (n°)	5
Microphones (n°)	3
Selection threshold	14

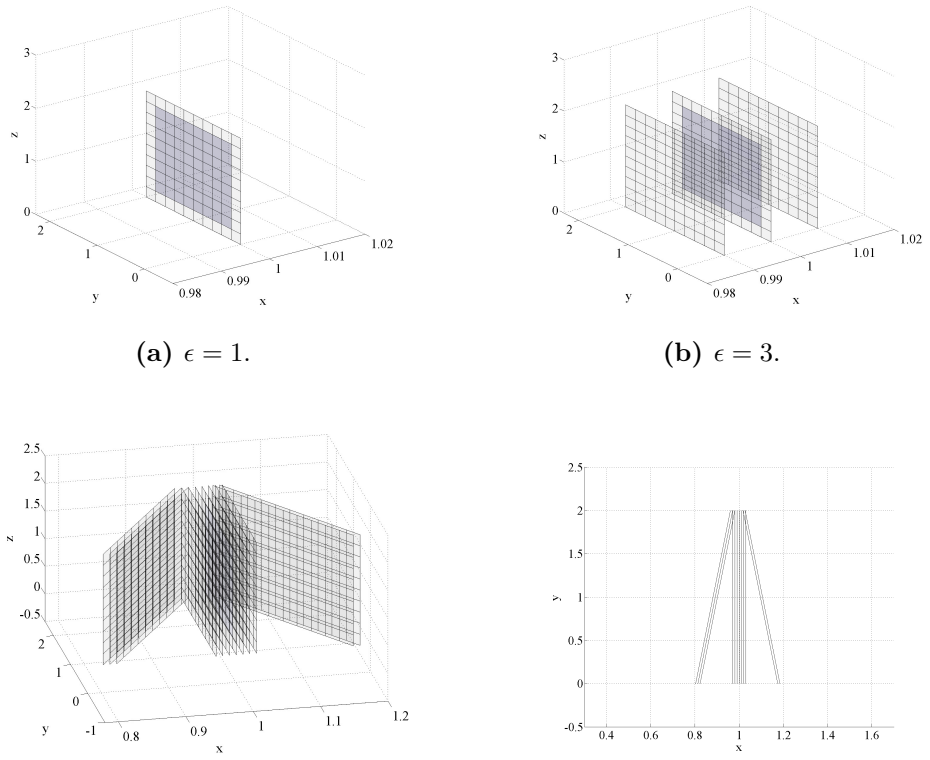
samples localized planes are three, and the actual reflector in the scene is that in in the middle of the group, lying above the darkest one: in fact, the three planes provides values of TOAs very similar one to each other, and it makes sense that increasing a little the acceptance threshold, there are included the nearest planes to the real one that provides the littlest differences from actual Times Of Arrival. Increasing again the threshold the algorithm begins to include other objects: in our case, it appears to localize three many regions, but actually decreasing the resolution of the x axis in the graph, we notice that all reflectors can be considered as belonging to the same cluster. This fact becomes more visible looking at the up view.

We report results in a more readable way in a summarizing table (Table 7.3). It includes the spherical coordinates of the plane to estimate, and those

Table 7.3: Increasing the acceptance threshold ϵ.

Threshold value	ρ[m], $\theta[^\circ]$, $\phi[^\circ]$	$\hat{\rho}$[m], $\hat{\theta}[^\circ]$, $\hat{\phi}[^\circ]$	$\lvert\mathbf{l} \cdot \hat{\mathbf{l}}\rvert / (\lvert\lvert\mathbf{l}\rvert\rvert \, \lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
$\epsilon = 1$	(1,0,0)	(1,0,0)	1
$\epsilon = 3$	(1,0,0)	(0.99,0,0)	1
$\epsilon = 8$	(1,0,0)	(0.98,0,0)	0.99

of the localized ones. Moreover, we provide a measure of the faithfulness of results, computing the inner product between the vectors parameter $\mathbf{l}$ referred to the real plane and $\hat{\mathbf{l}}$, relative to the estimated one. Dividing the product of these two terms by the same product with now the normalized



(a) $\epsilon = 1$.

(b) $\epsilon = 3$.

(c) $\epsilon = 8$. *Front view.*

(d) $\epsilon = 8$. *Top view.*

Figure 7.6: Several tests for increasing threshold ϵ.

vectors parameter, we obtain the cosine of the angle between vectors. This value is 1 when the reflectors are overlapped, and decreases if the estimate becomes worse. Obviously, to compute this measure, we have to obtain a single reflector from the provided estimate. We adopt the strategy whereby the selected plane in a cluster, it that for which the cost function assumes the minimum value.

Moreover, this chosen plane will be fed to the minimization task as starting point for the iterative algorithm.

7.1.2 Varying the size of the space parameter

In this simulation we want to test the algorithm depending on the number of values ρ, ϕ, θ that we feed in input, i.e. how much the space parameter appears as a dense distribution of values.

For our simulation we use an oblique plane. We create a space parameter

like that in Table 7.4. We want to test the algorithm trying different stepsize values, given a parameter. We choose the radial coordinate ρ, but it is the

Table 7.4: Space parameter $\{\phi, \theta, \rho\}$.

Parameter	Min value	Max value	Stepsize
ρ[m]	0.5	1.2	0.01,0.1,0.3
θ[°]	-90	90	5
ϕ[°]	0	355.5	4.5

same thing with the other two angular coordinates. Values of other involved variables are reported in Table 7.5.

Table 7.5: Involved parameters.

Parameter	Value
Sources (n°)	5
Microphones (n°)	3
Acceptance threshold	3
Selection threshold	14

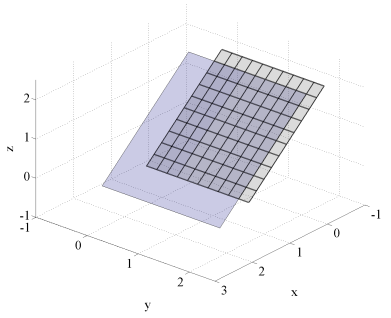
The results are shown in Figure 7.7 and summarized in Table 7.6. With values 0.1 and 0.01 the estimate does not substantially change and is faithful to the real geometry. The only difference is that with a smaller stepsize, more reflectors close to the real one may be included in the estimate. Increasing the stepsize until 0.3, we greatly decrease the number of fed values of ρ. In fact, the estimate is visibly different from the reflector present in the scene, since the Hough map does not contain the combination of parameters corresponding to its position. In many cases, the result may be not so lucky:

Table 7.6: Varying the space parameter.

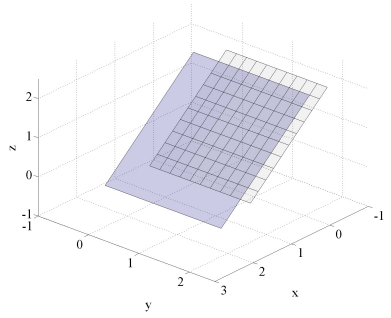
Stepsize value	ρ[m], θ[°], ϕ[°]	$\hat{\rho}$[m], $\hat{\theta}$[°], $\hat{\phi}$[°]	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
Stepsize = 0.01	(1,45,0)	(1,45,0)	1
Stepsize = 0.1	(1,45,0)	(1,45,0)	1
Stepsize = 0.3	(1,45,0)	(0.8,50,355.5)	0.991

working with so few values, there is the risk to estimate no plane.[1]

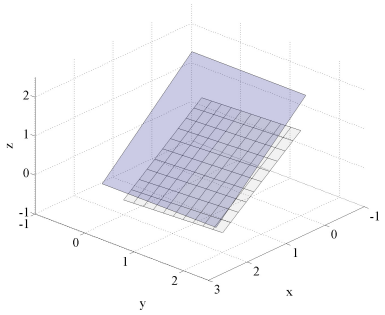
[1]Furthermore, we need to make lower the threshold through which cells voted are



(a) *stepsize* = 0.01(two overlapped planes).



(b) *stepsize* = 0.1.



(c) *stepsize* = 0.3.

Figure 7.7: Several tests for increasing stepsize.

7.1.3 Varying the selection threshold

As explained in previous chapters, another fundamental parameter that the Hough algorithm uses is the threshold identifying which points of the map extract and which ones discard. This parameter comes from that is called *hypotheses validation*, since it describes how much confirmations are needful to assess that the estimate is faithfulness. In practice, this threshold aims to select from the three-dimensional grid the more voted cells. These points are counted a number of times depending on how much TOAs are been found similar to those of the real reflectors.

Therefore, with next tests, we want to show how the behavior of the algorithm changes, along with the estimate, varying this parameter. We aim to give a practical confirmation to considerations given in Chapter 6,

extracted from the map: the stepsize is now too high and guessed planes will most likely be not so near to the real one. We discuss better this issue in next section.

concerning the problem of how, a suitable threshold, should be chosen.

We consider a single oblique plane in the scene. Involved parameters are set like in Table 7.7 and Table 7.8.

Table 7.7: Space parameter $\{\phi, \theta, \rho\}$.

Parameter	Min value	Max value	Stepsize
ρ[m]	0.9	1.1	0.01
$\theta[^\circ]$	-90	90	10
$\phi[^\circ]$	0	359	1

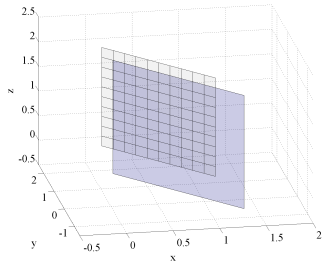
Table 7.8: Other parameters: sources, receivers, thresholds.

Parameter	Value
Sources (n°)	12
Microphones (n°)	3
Acceptance threshold	2

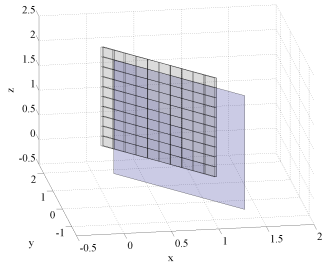
We aim to perform different tests progressively decreasing the threshold. We start selecting it to 35, since the number of acquisitions is 36. Results for each threshold value are shown in Figure 7.8. In two first tests the estimate is similar: with a threshold equal to 35 the localized plane matches perfectly to the real reflector, while setting it to 30 the algorithm gives three reflectors almost overlapped.

Performing other two tests with lower thresholds, as 20 and 10, we meet two almost identical situations: the algorithm localizes a rich cluster of planes, around the real one. The density of clusters is in inverse proportion to the value of the threshold.

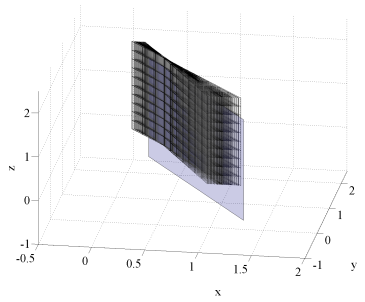
In the end, setting the threshold to 2, the result is so complicated that we had to reduce sizes of clusters just in order to show it graphically. The adopted strategy aims to incorporate all reflectors of the cluster in that plane on one at the end of it, according to the strategy described in Chapter 6. After this task, we obtain different main regions of concentration in space. A faithful estimate is represented by only one plane, although this result is still the worst among all tests just discussed, since it is the furthest from the true plane.



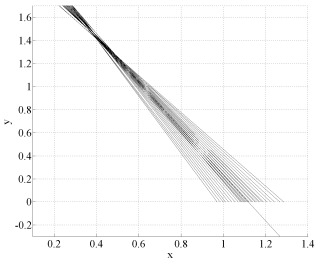
(a) *threshold* = 35.



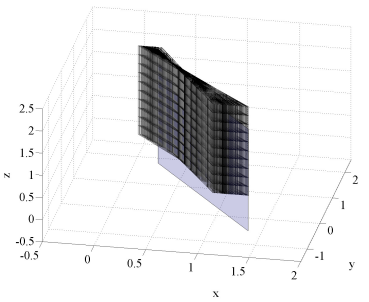
(b) *threshold* = 30.



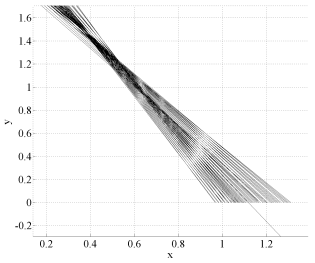
(c) *threshold* = 20.*Front view.*



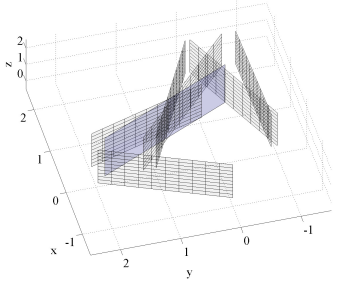
(d) *threshold* = 20.*Top view.*



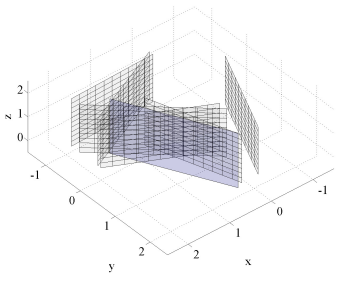
(e) *threshold* = 10.*Front view.*



(f) *threshold* = 10.*Top view.*



(g) *threshold* = 2.*Lateral view.*



(h) *threshold* = 2.*Front view.*

Figure 7.8: Several tests for decreasing selection threshold.

Summarizing, decreasing the threshold, the number of localized planes tends to grow, and in some cases the estimate becomes worse. Table 7.9 summarizes results.

In general, before any strategy for grouping, analyzing the density of clusters, it is possible to remove the wrong reflectors: their clusters are most likely the poorest ones.

Table 7.9: Varying the selection threshold.

Threshold value	ρ[m], $\theta[^\circ], \phi[^\circ]$	$\hat{\rho}$[m], $\hat{\theta}[^\circ], \hat{\phi}[^\circ]$	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
Threshold = 35	(1,0,26.5)	(1,0,26.5)	1
Threshold = 30	(1,0,26.5)	(0.99,0,26)	1
Threshold = 20	(1,0,26.5)	(0.96,0,24.5)	0.9995
Threshold = 10	(1,0,26.5)	(0.91,0,22.5)	0.9976
Threshold = 2	(1,0,26.5)	(0.9,0,18)	0.9928

7.2 Experiments with real data

In this section we show the results of some experiments with real data we performed both in dry room and in common room. Experiments that took place in a dry room are accomplished positioning some non-absorbent panels. In our case, we perform localization for one and two panels. The situation of more than two reflectors involved will be treated in experiments in real rooms. In some cases, performing the same test, we change the number of microphones or sources in order to check the robustness of the algorithm. However, these numbers have to be chosen depending on the kind and complexity of tests we accomplish.

A last importance note we have to do, is that unlike simulations, in which we "acquired" TOAs computing them by hand knowing positions of reflectors, sources and microphones, in real experiments, we obtain values of TOAs taking them from the impulse response, i.e. the cross-correlation between the acquired signals by microphones and the emitted one. Therefore, delay times are extracted as significant peaks of the impulse response and can include some form of noise. This fact brings to another not trivial concept: performing a simulation we consider loudspeakers and receivers as omnidirectional

objects. But in our real case, it is not so: in our experimental setup, microphones are omnidirectional, but sources are highly directive. This can have a great effect on the amplitude of peaks, that may not be homogeneous in all acquisitions, creating the problem of which should be the better strategy to take peaks.

In general, however, the issue of choosing a suitable threshold over which select maxima, and below which discard them, is not simple: there is the risk of extracting also noise if we use a too low threshold, or select too few peaks if it is chosen too high. It could be useful exploit clever techniques for the peaks extraction, but in practice the task of placing microphones and sources can help the issue: in many situations, doing it with special care and craftiness we also we can act just to detect desired information (detect an edge rather than a wall,etc).

In the first paragraph we describe the experimental setup. In the others, we propose experiments in anechoic and common rooms. We summarize results in a table, in which a comparison between the actual geometry and the estimated one is more readable.

7.2.1 Experimental setup

The hardware we use in our experiments includes one small loudspeaker and several microphones, obviously together with a sound card connected to a computer. The loudspeaker needs to be powered by an amplifier. We use only one source and we move it by hand in order to perform more tests and record multiple acquisitions. In some cases the source is fixed to a turntable allowing 24 positions at 15°. In other situations the loudspeaker is fixed on a horizontal support.

Moreover, this kind of loudspeaker is highly directive. This fact affects the way of choosing a suitable configuration of sources and receivers, respect to the reflectors in the environment. The directivity mainly influences the amplitudes of peaks recorded, which can differ a lot from an acquisition to another.

However, we return back to describe hardware. Receivers are *"Beyerdynamic MM1"* microphones whose specifications are described in Table 7.10. We use them from 4 until 6 for each experiment, depending on the kind of

test. They can be positioned randomly fixed on different bars, or fixed all together in array configuration. Main features of these microphone are:

- Linear Frequency response
- Omnidirectional Polar Pattern
- Calibrated Open Circuit Voltage
- Narrow Tubular Construction

Table 7.10: Experimental setup: microphone specifications.

Model	*Beyerdynamic MM1*
Operating principle	Pressure
Transducer type	Condenser (back electret)
Operating principle	Pressure
Polar pattern	Omnidirectional, diffuse field calibrated
Open circuit voltage at 1 kHz (0dB = 1V/Pa)	15 mV/Pa (= -36.5 dBV) ±1 dB
Nominal impedance	330 ohm
Load impedance	2.2 kohm
Connector	3-pin XLR
Length	133 mm
Shaft diameter	19 mm
Head diameter	9 mm
Weight without cable	88 g
Frequency response	20 - 20.000 Hz (50 - 16.000 Hz ±1.5 dB)
Max. SPL at 1 kHz	128 dB
S/N ratio rel. to 1 Pa	$>$ 57 dB
A-weighted equivalent SPL	approx. 28 dB(A)
Power supply	12 - 48 V phantom power
Current consumption	approx. 3.4 mA

Now, we say some words about the sound card, whose specification are widely described in Tables 7.11, 7.12, 7.13, 7.14, 7.15, 7.16. The model is

Table 7.11: Experimental setup: *Aurora Lynx 16* specifications (Analog I/O).

Analog I/O	
Aurora 16	Sixteen inputs and sixteen outputs
Type	Electronically balanced or unbalanced
Level	+4 dBu nominal / +20 dBu max. or -10 dBV nominal / +6 dBV max
Input Impedance	Balanced mode: 24 kohm Unbalanced mode: 12 kohm
Output Impedance	Balanced mode: 100 ohm Unbalanced mode: 50 ohm
Output Drive	600 ohm impedance, $0.2\mu F$ capacitance
A/D and D/A Type	24-bit multi-level, delta-sigma

"Aurora Lynx 16" and it is used together with the *"Focusrite Octopre LE"* analog-to-digital/digital-to-analog converters.

This soundcard allows several sampling frequencies: 44.1, 48, 88.2, 96, 176.4, 192 kHz. We work to 44.1 kHz, to check the effectiveness of the algorithm in the worst case.

Moreover, the soundcard latency affects the values of TOAs acquired from the acquisitions and in order to obtain the most faithful dataset of TOAs to use in the Hough transform, we have to subtract to each value, the samples relative to the latency. With the specific soundcard we use, this value is 32 samples.

Some images of loudspeakers and microphones we use in experiments are shown in Figure 7.9. The source is shown in possible configuration, that is fixed on a circular support in which the white lines colored on the circumference, indicates the 24 available positions. In this case it has to be moved by hand for each acquisition. But there exist other suitable configurations, depending on the experiment.

Table 7.12: Experimental setup: *Aurora Lynx 16* specifications (Analog In Performance).

Analog In Performance	
Frequency response	20 Hz - 20 kHz, +0/-0.1 dB
Dynamic range	117 dB, A-weighted
Channel crosstalk	-120 dB maximum, 1 kHz signal, -1 dBFS
THD + N	-108 dB (0.0004%) @ -1 DBFS -104 dB (0.0006%) @ -6 DBFS 1 kHz signal, 22 Hz - 22 kHz BW

Table 7.13: Experimental setup: *Aurora Lynx 16* specifications (Analog Out Performance).

Analog Out Performance	
Frequency response	20 Hz - 20 kHz, +0/-0.1 dB
Dynamic range	117 dB, A-weighted
Channel crosstalk	-120 dB maximum, 1 kHz signal, -1 dBFS
THD + N	-107 dB (0.0004%) @ -1 DBFS -106 dB (0.0006%) @ -6 DBFS 1 kHz signal, 22 Hz - 22 kHz BW

Table 7.14: Experimental setup: *Aurora Lynx 16* specifications (Digital I/O).

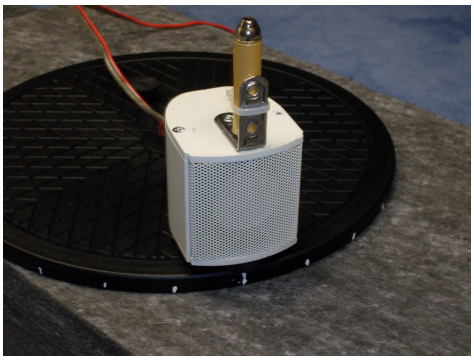
Digital I/O	
Number / Type	16 inputs and 16 outputs 24 bit AES/EBU format, transformer coupled
Channels	16 in/out in single-wire mode 8 in/out in dual-wire mode
Samples Rates	All standard rates and variable rates up to 192 kHz in both single-wire and dual-wire modes

Table 7.15: Experimental setup: *Aurora Lynx 16* specifications (On-board Digital Mixer).

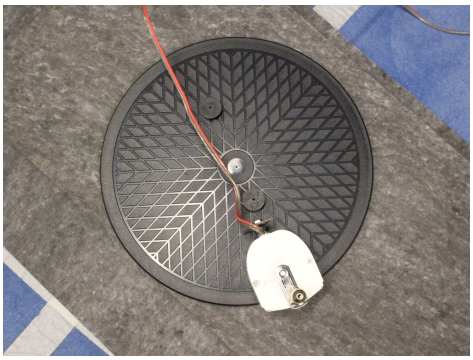
On-board Digital Mixer	
Type	Hardware-based, low latency
Routing	Ability to route any input to any or multiple outputs
Mixing	Up to four input or playback signals mixed to any output, 40-bit precision
Status	Peak levels to -114 dB on all inputs and outputs

Table 7.16: Experimental setup: *Aurora Lynx 16* specifications (Connections).

Connections	
Digital I/O Ports	25-pin female D-sub connectors. Port A: channels 1-8 I/O, Port B: channels 9-16 I/O. Yamaha pinout standard
Analog I/O Ports	25-pin female D-sub connectors. Analog In 1-8, Analog In 9-16, Analog Out, 1-8 Analog Out 9-16. Tascam pinout standard
External Clock	75-ohm BNC word clock input and output
MIDI	One input and one output. Standard opto-isolated, 5-pin female DIN connectors



(a) *Loudspeaker(zooming).*



(b) *Loudspeaker(turntable).*

(c) *Microphones(zooming).*

(d) *Microphones(configuration).*

Figure 7.9: Some examples of hardware.

7.2.2 Localization in dry room

We perform various tests in this situation in order to check the efficiency of the algorithm. Since walls (panels) have not infinite size, it is a crucial point the correct positioning of microphones and sources: we must be sure that a acoustic path will be established, and over that it corresponds to the reflected signal and not to a diffracted one. So, we must ensure that the reflection point lies on the panel, and not outside or on its edge. This issue may compromise the correct working of the algorithm.

For reasons of simplicity we start with a single non-absorbent panel in an anechoic room, initially in a vertical position (parallel to a wall) and then

variously inclined. Later we treat two panels, in different positions.

Single vertical plane

We place one loudspeaker on a vertical support and six microphones on a horizontal one, close one to each other but not in array configuration. The source is oriented toward the panel, because of its directivity. If we analyze the impulse response for the first pair source-microphone shown in Figure 7.10, we notice that it exhibits two relevant peaks: the first one corresponds to the direct signal, and the second one to the reflected signal. There is a third peak, but its amplitude is not significant. It may result from other small objects in the room.

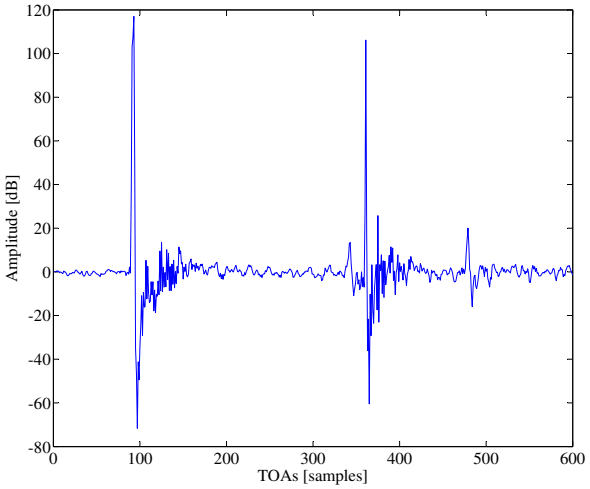
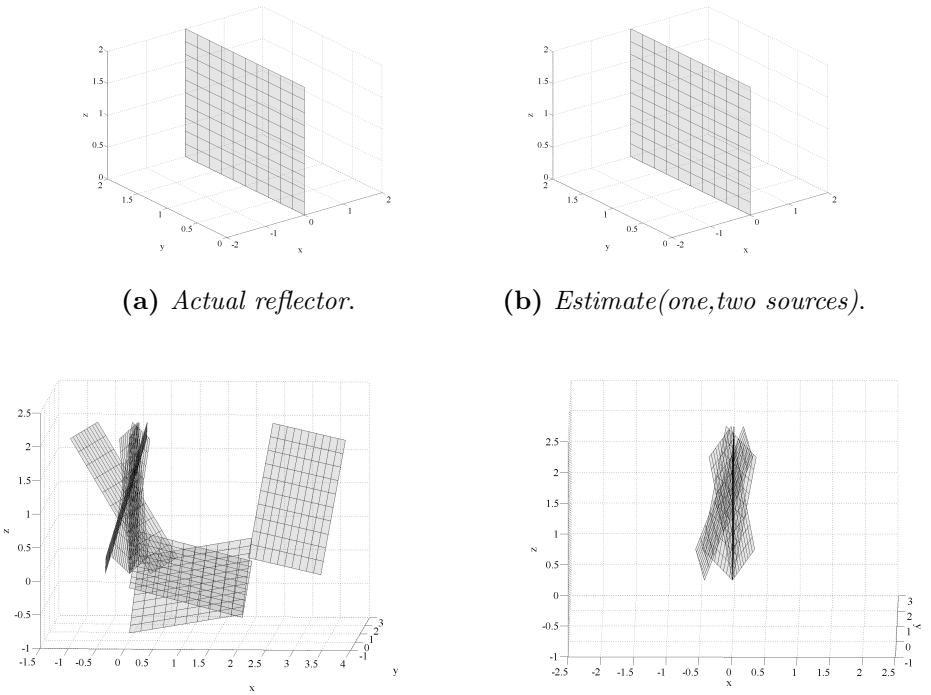


Figure 7.10: Cross-correlation between the emitted signal and the received one. It shows three peaks, of which the first two ones are relevant.

We test the algorithm before with one source and then two sources, adding a new position to the previous one, to have the double number of acquisitions. The results are equal and for this reason we include the relative graph just one time.

Increasing the number of sources has the main effect to improve the robustness. In these specific conditions this fact is not noticeable. But what happens if we decrease the threshold through which we take points from the Hough map? The results, together with the previous ones, are shown in

Figure 7.11. As we can notice, with more available data, the region of concentration of planes becomes smaller and does not include stranger objects. In order to provide readable results in Table 7.17 and dealing with clusters of estimated planes, we always consider a single plane, the closest to the actual one.



(a) *Actual reflector.* **(b)** *Estimate(one,two sources).*

(c) *Estimate(one source,low threshold).* **(d)** *Estimate(two sources,low threshold).*

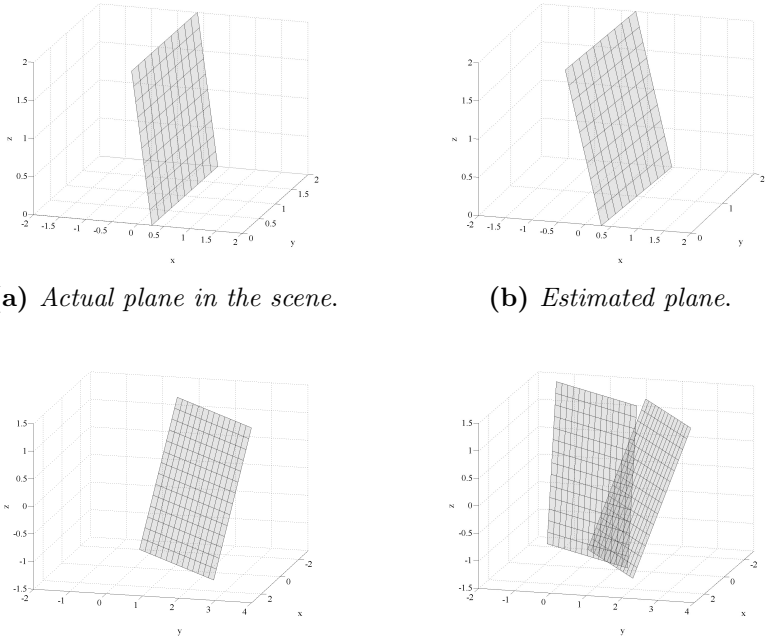
Figure 7.11: Localizing a single vertical plane with changing parameters.

Table 7.17: Localizing a single vertical plane with changing parameters.

Experiment	ρ[m], $\theta[^\circ], \phi[^\circ]$	$\hat{\rho}$[m], $\hat{\theta}[^\circ], \hat{\phi}[^\circ]$	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
1 source(high thresh.)	(0,0,0)	(0,0,0)	1
1 source(low thresh.)	(0,0,0)	(0.13,11.5,0)	0.98
2 sources(high thresh.)	(0,0,0)	(0,0,0)	1
2 sources(low thresh.)	(0,0,0)	(0.28,-14.3,0)	0.91

Single oblique plane

We place one loudspeaker and six microphones, using exactly the same positions of the previous test in the last subsection. We accomplish two experiments placing, in the first one, the panel obliquely, with one axis parallel to y axis, while in the second one a panel with no axis parallel to x, y or z. The results are shown in Figure 7.12. As you notice in the second situation, the algorithm provides two near planes. That with the minimum value of cost function is fed to the minimization task compared. The result is reported in Table 7.18, together with the others.



(a) *Actual plane in the scene.* **(b)** *Estimated plane.*

(c) *Actual plane in the scene.* **(d)** *Estimated plane.*

Figure 7.12: Localizing a single plane variously inclined.

Table 7.18: Localizing a single plane variously inclined.

Experiment	ρ[m], $\theta[^\circ], \phi[^\circ]$	$\hat{\rho}$[m], $\hat{\theta}[^\circ], \hat{\phi}[^\circ]$	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
Test 1	(0.26,11,0)	(0.3,20,355.5)	0.9934
Test 2	(1.18,-16.6,134.6)	(0.9,-15,144)	0.9881

Two not mutual planes

Now we move to localization of multiple planes. We start with two not mutual panels. The angle between them is obtuse. We use again one source and six microphones as before. In this situation the impulse response for each acquisition will be more complex than in the case with a single plane. More in detail, it has to contain multiple peaks relative to first-order reflections (two reflectors), in addition to the peak of the direct signal. An example of impulse response is shown in Figure 7.13. This is relative to the first pair source-receiver. As we can see, the impulse response exhibits various peaks,

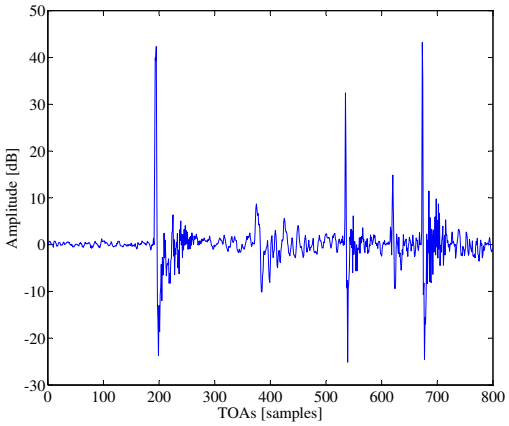


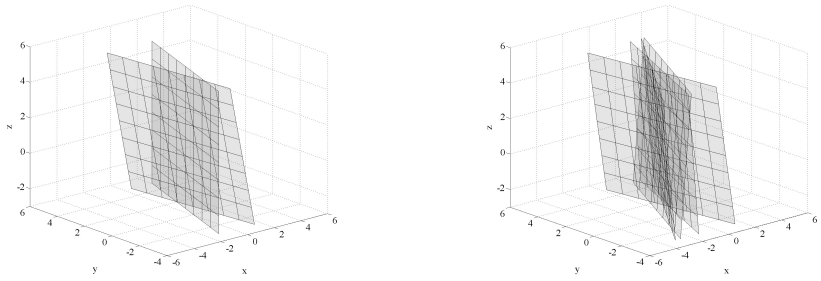
Figure 7.13: Impulse response. It shows three relevant peaks, of which the first one correspond to the direct signal, while the other two, to reflections.

of which three are relevant. The first one is relative to the direct signal, while the other two correspond to first order reflections on the two planes.

The algorithm detects four reflectors (Figure 7.14): three reflectors close one to each other and another plane apart from them. Each of two regions identified by the localization process, are faithful respect to the actual geometry of the environment. Again, data are compared after minimization.

Table 7.19: Localizing two not mutual planes.

Plane (n°)	ρ[m], θ[°], ϕ[°]	$\hat{\rho}$[m], $\hat{\theta}$[°], $\hat{\phi}$[°]	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
1	(0.51,9.3,6.7)	(0.5,10,9)	0.7646
2	(0,0,158.1)	(0.6,-5,139.4)	0.9986



(a) *Actual plane in the scene.* **(b)** *Estimated planes.*

Figure 7.14: Localizing two not mutual planes.

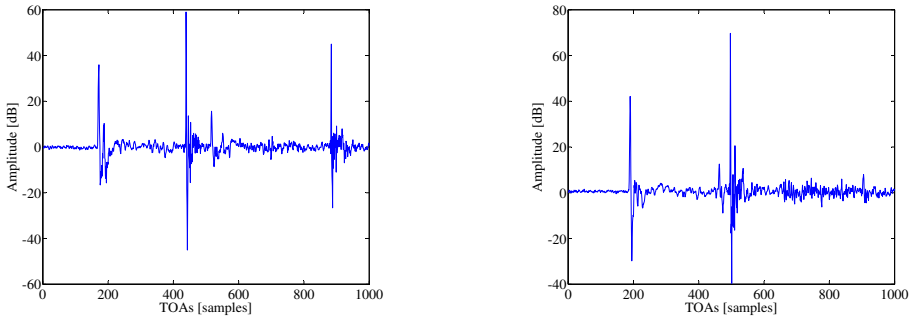
Two facing planes

Now we consider the case of facing planes. We treat two different situation: the first one in which the two panels are placed at 90°, the second one in which they are parallel.

The main difference from previous tests, is that dealing with facing planes there not are only first-order reflections, but also high-order signals, and also their contribute will appear in the impulse response. However, our algorithm has to perform localization entirely based on first-order reflections. Consequently, now the problem of directivity of loudspeakers assumes bigger importance. In fact, think what happens if a source is located midway between two parallel panels: if we direct the loudspeaker toward one of them, necessarily the two detected peaks differ a lot in amplitude. This fact can create ambiguity about the choice of which strategy exploit in order to pick maxima. If we work through a threshold, and if we set it low we risk to include also noise, while if we set it too high peaks of interest may be lost. A trick may be to exclude the direct signal picking maxima. This can be useful when its amplitude is very higher respect to other peaks.

We place the source in four positions. Microphones are again six with the same locations of previous tests. So we have 24 acquisitions. For simplicity, the loudspeaker is placed on a circular support with diameter equal to 30 cm, and with angle between positions of 15° (24 positions). Loudspeaker are placed in a way such that two sources for plane are highly directive toward it. For each pair, positions of sources are consecutive. Thus, we have ensured that peaks of interest for each panel (first-order reflections) have a consistent amplitude, so for sure they will be taken in account. We start with the case of

perpendicular panels. In Figure 7.15 we show two impulse responses relative respectively to the first and third source, directed to different planes.



(a) *Impulse response for a source directed toward the first plane.*

(b) *Impulse response for a source directed toward the second plane.*

Figure 7.15: Impulse responses for two facing reflectors. Both figures exhibit a peak for the direct signal and peaks relative to first order reflections on the two panels. Only one response shows a peak relative to a second-order reflection.

As we can notice, in addition to the peak of the direct signal, we find other maxima. In the figure on the left there are three peaks between 400 and 600 samples(they are better visible zooming the region). This happens because initially the experiment was been performed placing three panels, and so acquisitions include also peaks referred to the plane in addition. Peaks corresponding to our actual reflectors are the first two, because the third reflector was located at a larger distance from sources and receivers than the other panels. However we take in account only tracks relative to two planes of interest, so we discard those peaks. Moreover, it appears at more on less 900 samples another maxima. This is most likely relative to a second-order reflection.[2]

If now we take a look to the figure on the right, we find a similar situation. In fact, three peaks between 400 and 600 samples are still present, but dealing with the amplitude of the first two the situation is reversed (the second one is higher than the first one). This happens because of the directivity of loudspeakers: knowing that sources assumes more or less the same distance from planes, it is intuitive think that the maximum with the highest ampli-

[2]We could state it since sources are placed more on less at the same distances from three planes, so a peak with a very larger delay will probably not be a first-order reflection.

tude corresponds for sure to the plane toward which the source is directed, and the other peak refers to the remaining plane.

Furthermore in some case a second-order reflection is present. It depends on the disposition of panels and analysis through the image source method. This is why for some acquisitions we detect a peak at approximately 800-900 samples and in others it does not happen.

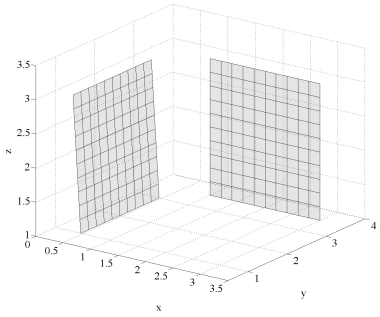
Now, dealing with the algorithm, in particular the Hough transform, we have to low the threshold through which we select points from the map. Although the number of acquisitions are 24, it is crucial remember previous considerations about disposition of sources and its directivity. In fact, we are interested to take in account for computation only relevant maxima in the impulse response, so peaks corresponding to the plane toward which sources are directed in that specific acquisition. In this way we constrain the algorithm to acquire not 24 TOAs (first-order reflections) for each plane but 12 (the number of more directive sources times the number of microphones). Hence, the threshold has to be set at most to 11. However, this restriction can be overcame involving omnidirectional objects.

After this needful explanations, we move to practical results. They are shown in Figure 7.16 and reported in Table 7.20.

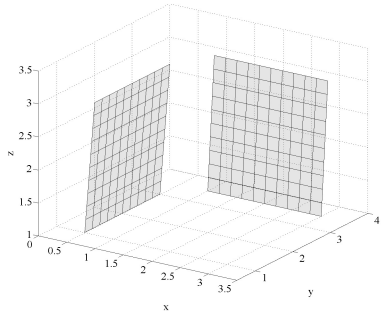
When, instead, we perform the test moving a panel, two positions of sources change, such that they are directive toward the new reflector. In this situation the planes are completely facing one to each other, and loudspeakers and microphones are midway between them, hence first-order reflections are probably related to similar time delays, while second-order reflections to very longer delays. The estimated results are joint in the figure and table together with those of the first configuration.

Table 7.20: Localizing two facing planes.

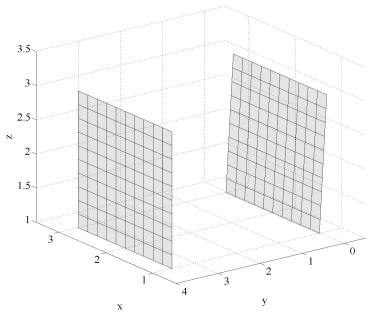
Experiment/Planes	ρ[m], θ[°], ϕ[°]	$\hat{\rho}$[m], $\hat{\theta}$[°], $\hat{\phi}$[°]	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
Test A - 1° plane	(3.545,0,90)	(3.4,-5,90)	0.9996
Test A - 2° plane	(0.53,4,0)	(0.4,-5,0)	0.9842
Test B - 1° plane	(3.545,0,90)	(3.4,-5,90)	0.9996
Test B - 2° plane	(0.13,5,90)	(0.1,0,90)	0.9958



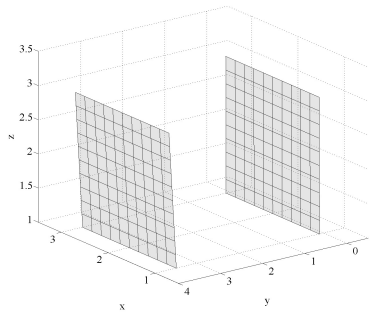
(a) *Actual geometry (1° configuration).*



(b) *Estimated reflectors (1° configuration).*



(c) *Actual geometry (2° configuration).*



(d) *Estimated reflectors (2° configuration).*

Figure 7.16: Localizing two facing planes.

7.2.3 Inference of a real environment

We conclude the series of experiments with a test in a common room. A situation like that can be complex, in fact we had to repeat tests more than one time in order to solve problems that came out in performing experiments.

The room has approximately sizes of 5 x 4 m^2, with height 3 m. Inside there are present some tables with chairs, and a cabinet. One wall is completely made of glass.

The advantage of a real situation is that the environment is closed, that is each wall border on another. Therefore we do not care so much about where reflection points lies, since if a ray does not pass through a wall it for sure crosses through a wall which shares a border with it, and thus a reflected ray is certainly detected. The only thing to avoid, if it is possible, may be to hit edges, that could give origin to diffracted paths, and this analysis is not included in our work.

Unfortunately there are also disadvantages. The main obvious drawback regards noise that can be detectable in a not anechoic room. It suffices to observe the impulse responses. Additionally, materials constituting obstacles affect how acoustic rays are reflected, because of their degree of absorption. In an anechoic room this does not happen, since we used the same kind of panel in the situation of multiple reflectors. In a real room, at the contrary, this problem can bring to not detect all needful peaks and relative TOAs, risking to lose some plane in the final estimate.

Furthermore, objects involved in the scene can affect the task of localization. It does not depend just on materials, but mainly on the fact that these obstacles can occlude acoustics paths or diffract sound. In our case, one way through which any additional material does not influence processing, is place microphones and sources in a clever way: tables and chairs bring no contribute if reflection points lie on a plane above them. At the contrary, the cabinet has to be taken in account, since it is almost as long as the whole wall, and so it is very likely that rays fall in this region.

In addition to these main drawbacks proper to a real situation, we encountered other problems coming from the non-depth knowledge of our experimental method in the three-dimensional case.

One problem we needed to solve results from the face made of glass. After various attempts in which that specific wall was not detected, analyzing involved impulse responses we noticed that peaks relative to first-order re-

flections had too low amplitude, being almost absent. Thus we gather that glass likely does not reflect in the best way acoustic rays. The solution we adopted consists of cover the wall with panels of wood, the same ones used in the anechoic room.

Another difficulty we overcame regards disposition of hardware in the environment. We had to accomplish several tests before finding the better geometry for hardware. Totally we placed the source, fixed on a vertical support, in twelve different positions, two for plane in order to ensure maximum available directivity for each wall. Even within a pair of sources, their coordinates differ a lot. Instead, microphones are four and fixed on vertical supports. They are placed more or less in the center of the room, each one of them on four corners of a rhomboid structure. Microphones differ mainly in height. These choices were analyzed with care.

The part of the algorithm concerning the Hough transform involves the parameters in Table 7.21.

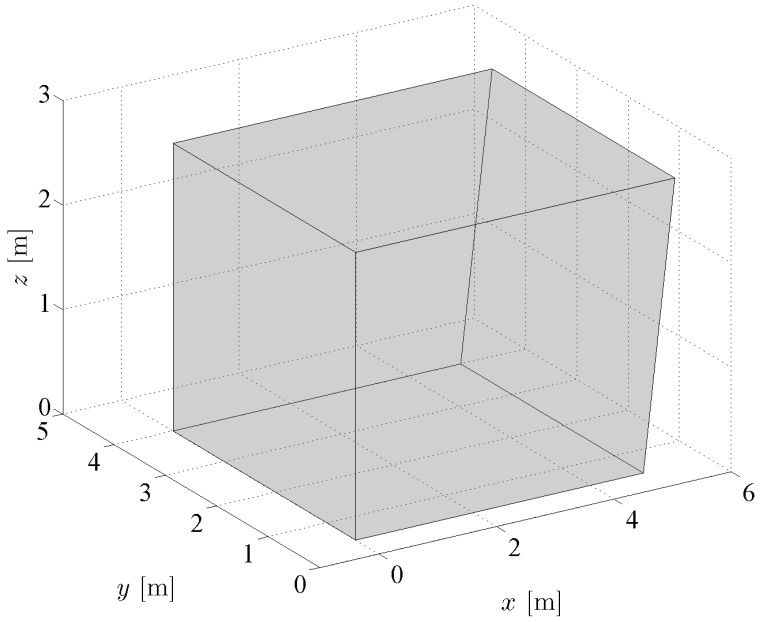
Table 7.21: Space parameter $\{\phi, \theta, \rho\}$.

Parameter	Min value	Max value	Stepsize
ρ[m]	0	5	0.01
θ[°]	-90	90	10
ϕ[°]	0	355.5	4.5

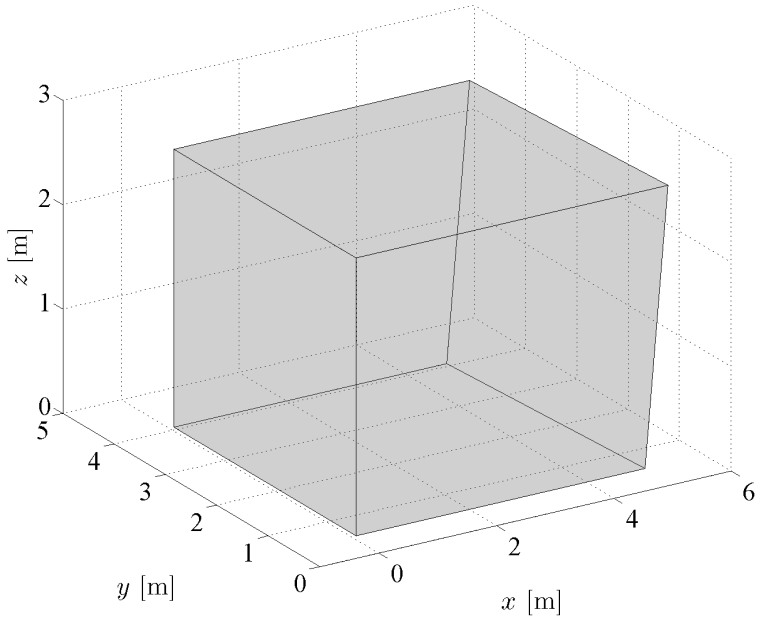
The threshold of selection from the Hough map is set to 7, since the number of acquisitions (TOAs) for each plane is 8. Now we look at results in Figure 7.17. The localization provides a good result. A more detailed comparison is shown in Table 7.22.

Table 7.22: Inference of a real environment.

Plane (n°)	ρ[m], θ[°], ϕ[°]	$\hat{\rho}$[m], $\hat{\theta}$[°], $\hat{\phi}$[°]	$\lvert\mathbf{l}\cdot\hat{\mathbf{l}}\rvert/(\lvert\lvert\mathbf{l}\rvert\rvert\,\lvert\lvert\hat{\mathbf{l}}\rvert\rvert)$
1	(0,0,0)	(0.03,0,0)	1
2	(4.8,-11,0)	(4.81,-9.99,4.5)	1
3	(0.45,0,90)	(0.48,0,90)	1
4	(4,0,90)	(4.02,0,90)	1
5	(0,90,0)	(0.02,90,0)	1
6	(2.75,90,0)	(2.68,90,0)	0.99



(a) *Geometry of the environment.*



(b) *Estimated geometry.*

Figure 7.17: Inference of a real environment.

Moreover, we can notice that the algorithm detects a reflector referred to the cabinet. If we wanted to estimate also the wall behind it, we should place sources and receivers in a way such that some reflections points lie on that wall, in the region not covered by cabinet.

Chapter 8

Future directions

The theoretical and experimental results we obtained allow us to draw some conclusions on the research done so far and to guess what could be the next steps. In the following we provide a list of some ideas for future developments.

- Proposed simulations and experiments were done using Matlab implementation of the algorithm, also because this language lends itself to a mathematical analysis, better than other languages. Nonetheless, unlike what happens with algorithm simulations and validations, in regard to real time applications the overall computation time is still unsuitable. Therefore, an improvement could be try to develop a more optimized and efficient implementation.

- Even if for real time applications the overall computation time is too high, after a suitable optimization of the algorithm, the technique with a continuously moving loudspeaker may allow to perform very fast acquisitions, in addition to the possibility of realtime visualization.

- Another issue that looks interesting to investigate is if it could be useful to process the different frequencies separately. In fact, through a frequency analysis on the reflected rays, i.e. on involved wavelengths, we can infer the extension of reflectors. This is an interesting improvement, since our algorithm allows to localize planes on which walls lie, and not faces with vertices. Thus, this task can provide a useful additional information.

In addition to these viable works, an important step is the implementation of diffraction phenomena. In this chapter we mention the main theoretical

concepts of diffraction theory and we propose a model formulation for the problem of diffraction by edges, together with an idea of solution. We shortly describe the main laws and how we could take them into account in computation, with the goal to propose a starting point for more detailed future works.

In the next section we enunciate key principles which diffraction theory is based on. In the others, we provide an idea of model formulation and problem solution.

8.1 Theory

The phenomenon of diffraction is associated to deviation of trajectories of rays when they meet obstacles along their path. Its analysis can be introduced and described by different theories.

One of these refers to the *Huygens's principle*, whereby each point of a wavefront can be considered as a secondary source of spherical waves. The wavefield in a point in space can be obtained as superposition of all secondary spherical wavelets that reach that point. Fresnel adds interference between waves to Huygens' theory. The diffracted field is obtained as the sum of contributions of secondary point sources, through which we discretize the edge.

Another theory recalls the *Helmoltz-Kirchoff* integral. It expresses the diffracted field as a function of the field on the surface of the diffracting object. Thus, it can be represented by a surface integral.

The geometric optics itself, would not be compatible with the phenomenon of diffraction. Nonetheless, there exist several techniques involving optics which can describe the diffraction theory.

The first extension of geometrical optics to take in account the diffracted field is the Joseph Keller's geometrical theory of diffraction (GTD). It exploits principles of geometrical optics including also diffraction, i.e. introducing diffracted rays in addition to the usual rays. The method is so based on three kinds of ray: direct, reflected and diffracted. Currently the more used theory is an extension of Geometrical Theory of Diffraction, named UTD (Uniform Theory of Diffraction). It an high frequency method for solving electromagnetic scattering problems and it applies to infinite edges.

Keller suggested that diffracted rays exist whenever a ray (directed or reflected) hits edges, corners, or vertices of boundary surfaces, or a surface arises irregularities. Moreover, Keller postulated diffraction theory starting from the generalized Fermat's principle: a beam diffracted by an edge between S and P follows the curve with the minimal optical path among all curves between S and P who have a point on edge. Therefore, the **law of diffraction** states that *the diffracted ray and the incident one lie on opposite sides with respect the orthogonal plane to the edge and passing through the diffraction point and angles these rays create with the edge are given by "Snell's law for diffraction"*:

$$\eta_i \sin\theta_i = \eta_d \sin\theta_d$$

If rays propagate in the same material $\theta_d = \theta_i$.

Keller noticed that in the case of oblique incident ray on the edge, diffracted rays propagated along cones having the edge as axis. The cone of diffracted rays has been named the *Keller's cone* and its aperture is equal to the angle of incidence. For each incident ray on a point of the edge, there are infinite diffracted ones belonging to the lateral surface of the Keller's cone, as shown in Figure 8.1.

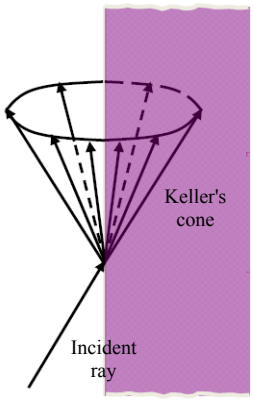
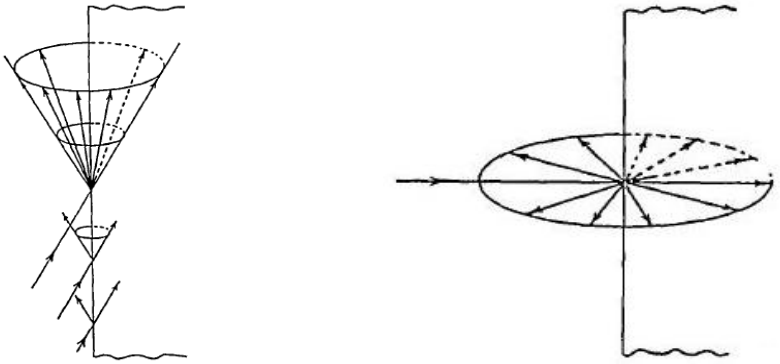


Figure 8.1: For each ray which hits a point on the edge, there are infinite diffracted ones belonging to the lateral surface of the cone. The angle of incidence corresponds to bevel of the cone.

Moreover, if the angle of incidence $\theta_i = 90°$ the incident beam is orthogonal to the edge and the diffraction cone degenerates into a disk, as we see in Figure 8.2. By definition, in 2D case, all rays are orthogonal to the edge.



(a) *The cone of diffracted rays produced by an incident ray which hits the edge obliquely.*

(b) *The plane of diffracted rays produced by a ray normally incident on the edge.*

Figure 8.2: Keller's cone. When the angle of incidence is equal to 90° the cone degenerates in a disk.

Thus, the diffracted wave is cylindric for normal incidence, conic for oblique incidence. However, for a given point source and point receiver location, the diffraction of a wave over an edge is represented by a single ray.

8.2 A model formulation

Analyze the behavior of edges and how they diffract sound, can result more complex respect to study how rays reflect onto planes with infinite extension. In fact, many concepts involved in the problem description and methods typical of the situation of specular reflection, cannot be extended to describe diffraction.

For our analysis we consider only infinitely long edges, so we analyze cases in which a diffracted ray is always established. A first limitation is that the image source method cannot be exploited, since dealing with the phenomenon of diffraction no image source exist. Moreover, there are infinite planes on which the edge can lie, so this adds ambiguity: an edge is identified by a line, thus there exists an infinite number of planes that contains it. According to these considerations TOAs of diffracted paths can not be computed exploiting the image source method. Thus, it is needed to know the position of diffraction point to obtain the value of time delay by summing

two distances: one from the source to the diffraction point, and the other from the the diffraction point to the receiver. Finally, we need to guess some strategies.

As explained in the previous section, the starting point for any consideration is the constraint we introduced before: the solid angle formed by the incident ray and the edge, in the diffraction point, is equal to that formed by the edge and the diffracted ray. Through this information and imposing that the sum of distances from source to edge and from edge to receiver is equal to the value of TOA of the diffracted path, we constrain potential diffraction points to lie on an ellipsoid with the source and the receiver as foci. If we now change positions of foci, referring again to the same edge, we obtain a different ellipsoid and a different diffraction point. Thus, fixing an edge, and moving source and receiver locations, all diffraction points are aligned on the edge and the line on which it lies can be estimated as the common tangent line to each ellipsoid. An illustration is shown in Figure 8.3.

However, if we are interested to compute diffraction points, it does not make sense to perform an intersection of all ellipsoids, because the position of diffraction points changes whenever foci move. Nevertheless, an idea for a

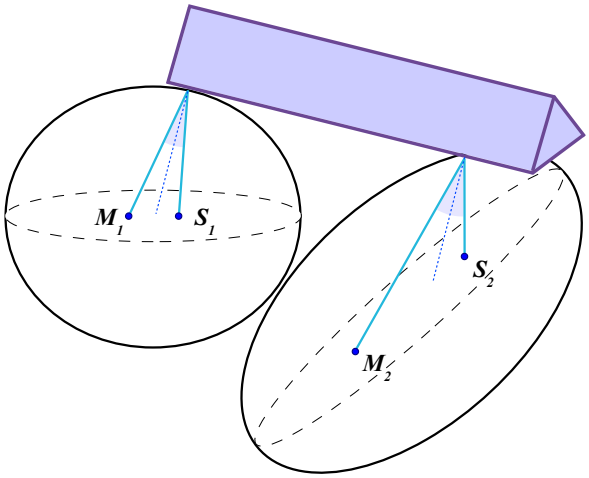


Figure 8.3: Given an edge, for different source and receiver locations, all diffraction points are aligned on the edge. Each diffraction point lies on an ellipsoid with foci the source and the receiver.

solution could that of finding the common tangent line to the ellipsoids, in order to compute the vector parameter representing the equation of the line containing the edge. This would help the issue of get an idea about extensions

of reflectors: by using the Hough Transform, we can consider and estimate planes with infinite extension. The issue of identifying an edge can "assign" a size to the reflector, providing an additional important information.

8.3 A proposed solution

In this section we want to propose a method to determine the diffraction point location on an edge, given positions of one source and one microphone.

Indicate with $\boldsymbol{S}$ the source position and with $\boldsymbol{R}$ the receiver one. The diffraction point is identified by $\boldsymbol{P}$. Moreover, take a unitary vector $\hat{e}$ tangent to the edge. Now, consider the incident ray from $\boldsymbol{S}$ to $\boldsymbol{P}$ and the diffracted ray from $\boldsymbol{P}$ to $\boldsymbol{R}$. As already explained, we operate imposing that the edge forms with the incident ray an angle equal to that formed with the diffracted ray. Indicates with $\hat{s} = \frac{\boldsymbol{P}-\boldsymbol{S}}{|\boldsymbol{P}-\boldsymbol{S}|}$ the normalized direction vector of the incident ray and with $\hat{r} = \frac{\boldsymbol{R}-\boldsymbol{P}}{|\boldsymbol{R}-\boldsymbol{P}|}$ the normalized direction vector of the diffracted ray. A representation is shown in Figure 8.4. Thus, we can write the constraint

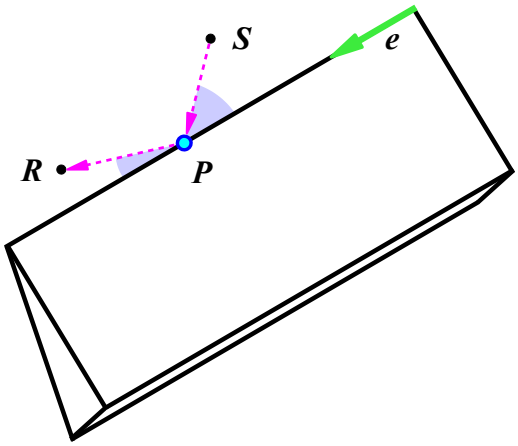


Figure 8.4: The diffraction point $\boldsymbol{P}$ is obtained through the equal angle constraint at the edge $\boldsymbol{e}$.

on angles, imposing the equality between inner products:

$$(-\hat{s})\hat{e} = \hat{r}(-\hat{e}) \tag{8.1}$$

In fact, since $\hat{s}$ and $\hat{e}$ are normalized vectors, inner products are equal to $\cos\theta_i$ or $\cos\theta_d$. Now our unknown is the point $\boldsymbol{P}$. This strategy holds also in the

situation of multiple diffractions between the source and the receiver. In this case we have to deal with a system containing as equations as the number of paths, in which the unitary vector $\hat{e}_i$ would be referred to the i-th edge and the normalized direction vector $\boldsymbol{P}_{i+1} - \boldsymbol{P}_i$ indicates the ray connecting two adjacent points. If we now parameterize the edge, the diffraction point can be rewritten as:

$$\boldsymbol{P}_i = \boldsymbol{O}_i + t_i \hat{e}_i$$

with $\boldsymbol{O}_i$ a reference point on the edge. So the system of equations can be rewritten in terms of unknown t_i. Then it is possible to compute diffraction points $\boldsymbol{P}_i$.

Chapter 9

Conclusions

In this thesis we have analyzed the scene reconstruction problem in the three-dimensional case. Our goal was that of extending all known mathematical concepts of 2D case to 3D one, and concerning the practical field, understand as how the algorithm had to change, as to learn new notions about techniques to perform in three dimensions, typical of this specific situation.

We described a method for the acoustic scene reconstruction problem that uses minimal hardware, discussing deeply theory on which it is based. Just for reason of complexity, in different steps in our work we proposed an heuristic approach as solution, and in many situations this strategy gave good results. Nevertheless, the proposed method can be integrated into other frameworks, providing greater results. The most severe limitations come from the constraint of working in the audible range, which leads to specular reflections. But even within these limits our approach can be improved.

In recent years the acoustic simulation in virtual environments found different fields of application. The most important application of the acoustic scene reconstruction is its use as the support tool for the sound field rendering. In order to simulate a virtual environment in a real one, we need to know approximately which is its geometry. In fact, this is the reason for which we do not consider an estimate as wrong if it detects clusters of planes in regions of actual ones. We are interested to get a sense of positions of obstacles. Moreover, our experiments were performed in empty environment, that is constituted only by walls or panels (in anechoic room) or objects placed in a way such that they did not influence the analysis. This choice has two reasons. First, localization of a three-dimensional geometry is a not so trivial issue, thus presence of other objects would make computation too

complex. Furthermore, but more crucial concept, the image source method, at the base of these applications, is inefficient in complex environments with a number of occlusions. Consequently, in this situation other kinds of techniques could help improve the performance.

We already mentioned other future developments in Chapter 8. As already said, the most important viable work is the implementation of diffraction phenomena, not covered by the current version of the algorithm. We only included some theoretical concepts and an idea of solution, because a deeper treatment does not fall within our analysis. The most important result we discussed, is that an analysis on diffraction by edges can give us information about the extension of reflectors, making us able to find the line in space which contains the edge and so "cutting" in that point the plane. This is an interesting improvement, since in this thesis, dealing with reflection, we cannot get any kind of information about sizes of objects. Therefore, also for these reasons, diffraction turns to be very crucial for an accurate simulation of an acoustic environment, thus it should to be an interesting issue for future developments.

Bibliography

[1] V. Murino and A. Trucco, "Three-dimensional image generation and processing in underwater acoustic vision ", *Proceedings of the IEEE Trans. in Audio, Speech and Language Processing*, vol. 88, december 2000, pp. 1903 - 1948.

[2] U. Castellani, A. Fusiello, V. Murino, L. Papaleo, E. Puppo, S. Repetto and M. Pittore, "Efficient on-line mosaicing from 3D acoustical images ", in *OCEAN 04.MTTS/IEEE TECHNO-OCEAN 04*, vol. 2, march 2005, pp. 670 - 677.

[3] U. Castellani, A. Fusiello, V. Murino, L. Papaleo, E. Puppo and M. Pittore, "A complete system for on-line 3D modelling from acoustic images ", *Signal processing. Image communication*, vol. 20, october 2005, pp. 832 - 852.

[4] V. Murino, A. Trucco and C. S. Regazzoni, "A probabilistic approach to the coupled reconstruction and restoration of underwater acoustic images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, january 1998, pp. 9 - 22.

[5] H. Sekkati and S. Negahdaripour, "Direct and indirect 3-D reconstruction from opti-acoustic stereo imaging",in *Third International Symposium on 3D Data Processing, Visualization and Transmission*, june 2006, pp. 615 - 622.

[6] S. Negahdaripour, H. Sekkati and H. Pirsiavash, "Opti-acoustic stereo imaging, system calibration and 3D reconstruction", *IEEE Conference on Computer Vision and Pattern Recognition, 2007*, june 2007, pp. 1 - 8.

[7] N. Hurts, X.Cuf, J.Salvi, "Integration of optical and acoustic sensors for 3D underwater scene reconstruction", in *Instrumentation Viewpoint*, Department of Computer Engineering, University of Girona, Spain, 2009.

[8] P. Krammer and H. Schweinzer, "Localization of object edges in arbitrary spatial positions based on ultrasonic data", *Sensors Journal, IEEE*, february 2006, pp. 203 - 210.

[9] J.A. Jimenez, M. Mazo, J. Urena, A. Hernandez, F. Alvarez, J.J. Garcia and E. Santiso, "Using pca in time-of-flight vectors for reflector recognition and 3D localization", *IEEE Trans. on Robotics*, october 2005, pp. 909 - 924.

[10] A.M. Sabatini, "Processing of specular echoes from planar reflectors in airborne sonar data", *Electronics Letters*, april 1993, pp. 699 - 700.

[11] A. ODonovan, R. Duraiswami and D. Zotkin, "Imaging concert hall acoustics using visual and audio cameras", *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 08*, april 2008, pp. 5284 - 5287.

[12] A. ODonovan, R. Duraiswami and J. Neumann, "Microphone arrays as generalized cameras for integrated audio visual processing", *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR 07*, june 2007, pp. 1 - 8.

[13] J. Meyer and G. Elko, "Spherical harmonic modal beamforming for an augmented circular microphone array", *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 08*, may 2008, pp. 5280 - 5283.

[14] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, vol. 1, 1999, pp. 666 - 673.

[15] T. J. Mckinley, M. M. McWaters and V. K. Jain, "3D reconstruction from a stereo pair without the knowledge of intrinsic or extrinsic parameters", *Proceedings of the Second International Workshop on Digital and Computational Video, DCV 01*, IEEE Computer Society, Washington, DC, USA, 2001, pp. 148.

[16] A.A. Montenegro, P.C.P. Carvalho and L. Velho, "Space carving with a hand-held camera", *Proceedings. 17th Brazilian Symposium on Computer Graphics and Image Processing, 2004*, october 2004, pp. 396 - 403.

[17] S. Se and P. Jasiobedzki, "Instant scene modeler for crime scene reconstruction", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2005. CVPRW 05*, june 2005, pp. 123.

[18] D. Lanman, D. Crispell, M. Wachs and G. Taubin, "Spherical catadioptric arrays: construction, multi-view geometry, and calibration", *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, june 2006, pp. 81 - 88.

[19] M. Kanbara, N. Ukita, M. Kidode and N.Yokoya, "3D scene reconstruction from reflection images in a spherical mirror", *Proceedings of the 18th International Conference on Pattern Recognition. ICPR 06*, IEEE Computer Society, Washington, DC, USA, 2006, pp. 874 - 879.

[20] N. Ragot, R. Rossi, X. Savatier, J.Y. Ertaud and B. Mazari, "3D volumetric reconstruction with a catadioptric stereovision sensor", *IEEE International Symposium on Industrial Electronics, 2008. ISIE 08*, july 2008, pp. 1306 - 1311.

[21] R. Orghidan, J. Salvi and E.M. Mouaddib, "Calibration of a structured light-based stereo catadioptric sensor", *Conference on Computer Vision and Pattern Recognition Workshop, 2003*, vol. 7, july 2003, pp. 70.

[22] J. Starck and A. Hilton, "Model-based multiple view reconstruction of people", *Proceedings of the Ninth IEEE International Conference on Computer Vision. CICCV 03*, IEEE Computer Society, Washington, DC, USA, 2003, pp. 915.

[23] L. Zhang, B. Curless and S.M. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming", *In The 1st IEEE International Symposium on 3D Data Processing, Visualization, and Transmission*, june 2002, pp. 24 - 36.

[24] A. Glassner, R. Cook, E. Haines, P. Hanrahan, P. Heckbert, and L. Speer, *An Introduction to Ray Tracing*, 4th ed. London, U.K.: Academic, 1987.

[25] R. Hartley and A. Zisserman, *"Multiple View Geometry in Computer Vision"*, Second Edition, Cambridge University, 2003.

[26] F. Antonacci, "Estimation of the geometric configuration of reflectors by multiple acquisitions", *IEEE Trans. in Audio, Speech and Language Processing*, vol. 10, no. 10, january 2009.

[27] F. Antonacci, A. Sarti and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions", *accepted for publication in IEEE Trans. in Audio, Speech and Language Processing.*

[28] A. Canclini, F. Antonacci, A. Sarti and S. Tubaro, "Acoustic source localization based on hyperbolic constraints in the homogeneous space", *IEEE Trans. in Audio, Speech and Language Processing*, vol. 6, no. 1, january 2007.

[29] A. Sarti, "SCENIC - Self-Configuring ENvironment-aware Intelligent aCoustic sensing", *ICT-FET*, december 2009.

[30] N. Tsingos, T. Funkhouser, A. Ngan and I. Carlbom, *Modeling Acoustics in Virtual Environments Using the Uniform Theory of Diffraction*, Princeton University.

[31] D. T. Blackstock, *Fundamentals of physical acoustic*, Austin University.

[32] L. L. Beranek, *Acoustics*, Cambridge University, 1993.

[33] C. Piersanti and F. Fuschini, *Teoria Geometrica della Diffrazione*, Università di Bologna.

[34] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves", *Comm. ACM*, vol. 15, january 1972.

[35] P.V.C. Hough, "Machine Analysis of Bubble Chamber Pictures", *Proc. Int. Conf. High Energy Accelerators and Instrumentation*, 1959.