POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Master of Science in
Computer Engineering

Multiple-Camera Plenacoustic Imaging: Estimation of the Radiation Pattern of an Acoustic Source

Candidate

Stefano Bider

Student Id. number 766377

Thesis Supervisor

Prof. Fabio Antonacci

Assistant Supervisor

Dr. Dejan Marković

Academic Year 2012/2013

POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Laurea Magistrale in
Ingegneria Informatica

Imaging Plenacustico Multi Camera: Stima del Pattern di Radianza di una Sorgente Acustica

Candidato

Stefano Bider
Matricola 766377

Relatore

Prof. Fabio Antonacci

Correlatore

Dr. Dejan Marković

Anno Accademico 2012/2013

Multiple-Camera Plenacoustic Imaging: Estimation of the Radiation Pattern of an Acoustic Source

Master thesis. Politecnico di Milano

This thesis has been typeset by LaTeX and the smcthesis class.

Author's email: stefano.bider@gmail.com

Sommario

Gli strumenti musicali acustici sono caratterizzati dallo specifico modo con cui interagiscono con l'ambiente circostante. Diversi studi sono stati e sono tutt'ora fatti sulla natura di questa interazione, su come questa si traduce in termini musicali, fisici e percettivi. Lo scopo di questo elaborato è definire un sistema che stimi il pattern di radianza di uno strumento, che è la funzione che descrive la dipendenza angolare dell'energia che questo irradia, e che costituisce uno degli aspetti più importanti della sua caratteristica interazione con l'ambiente. In particolare, il nostro scopo è progettare un sistema che lavori con più camere plenacustiche, dispositivi composti da una griglia di microfoni sincronizzati in grado, per mezzo di una serie di strumenti analitici, di misurare il campo sonoro, parametrizzarlo e determinare le caratteristiche della scena acustica.

Prendiamo come punto di partenza un sistema sviluppato nel 2013 presso l'Image and Sound Processing Group del Politecnico di Milano: in quel lavoro, la stima è effettuata usando una singola camera plenacoustica e un sistema separato per stimare la posizione nello spazio e l'orientamento dello strumento mentre viene suonato da un musicista.

L'idea fondamentale del nostro lavoro è che la tecnica può essere migliorata aggiungendo una seconda camera, in modo tale da permettere al musicista di muoversi in maniera più naturale e rendere più veloce il processo di acquisizione. Con questa tesi intendiamo esplorare più a fondo le potenzialità dell'analisi plenacustica, studiando come l'aggiunta di una seconda camera influenza le tecniche plenacustiche esistenti, e come l'interazione fra più camere può essere sfruttata per progettare un sistema autosufficiente, in grado sia di misurare il pattern di radianza che di stimare la posizione dello strumento. Presentiamo l'estensione formale al caso multi camera delle esistenti tecniche plenacustiche a camera singola, così come le tecniche di localizzazione, calibrazione e stima del pattern di radianza che abbiamo studiato per affrontare il problema in analisi.

Abstract

Acoustic musical instruments are characterized by their own specific way of interacting with the sourrounding environment. A variety of different studies have been and are still being made on the nature of this interaction, on how it translates into musical, physical and perceptual terms. The purpose of this work is to devise a system that estimates the radiation pattern of an instrument, which is the function that describes the angular dependence of the energy it radiates, and constitutes one of the most important aspects of its peculiar interaction with the environment. In particular, we aim at designing a system that relies on multiple Plenacoustic Cameras, devices composed by a set of synchronized microphones that are capable, by means of a series of analytic tools, of measuring the soundfield, parameterize it and assess the features of the acoustic scene.

We take as a starting point a system developed in 2013 at the Image and Sound Processing Group of Politecnico di Milano: in that work, the estimation is performed by using a single plenacoustic camera and a separate system for tracking the instrument position and orientation in space while it is played by a musician.

The founding idea of our work is that the technique can be improved by adding a second camera, in order to allow the musician to move in a more natural way as well as to shorten the acquisition process. With this thesis we intend to further explore the potentialities of plenacoustic analysis and study how the addition of a second camera impacts on the existing plenacoustic techniques, and how the interaction between multiple cameras can be exploited to design a self-contained system that is capable of both measuring the radiation pattern and estimate the instrument position in space. We present the formal extension of the existing single-camera pleancoustic techniques to the multiple-camera scenario, as well as the localization, calibration and radiation pattern measurement techniques that we have devised to address the problem at hand.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the most fascinating aspects of acustic musical instruments is that they exist to interact. They interact with people and with the environment in a very specific, complex way, and yet this interaraction is somehow transformed and rendered accessible to anyone; it is translated into a language that everyone is capable of understanding, that everyone perceives in a personal yet universal manner. Studying and modeling this interaction means exploring a whole universe, both from the musical and scientific points of view; determining the characteristics of an instrument means being able to devise ways to best record and amplify it, to assess what kind of environments are best suited to accommodate a live performance or a recording session. It can also mean being able to reproduce these characteristics, integrate them into synthesis and rendering systems in order to emulate the behaviour of real instruments in different contexts. Furthermore, a low level analysis of these characteristics can be brought into the manufacture process, which can be engineered in order to match the needs and requiremens of musicians and makers.

The aspects related to these studies are manifold, and cross over a variety of different disciplines. In this work, we focus our attention on one of the many possibile studies: the estimation of the radiation pattern of acoustic instruments. The radiation pattern is the description of the way in which the instrument radiates in all directions the environment in which it is immersed. This pattern can be very regular for some kind of sources, but when dealing with musical instruments it generally changes significantly as a function of the direction and of the frequency. The radiation pattern estimation problem has been addressed in the literature several times, with an incredible variety of approaches that depend on the intent of the studies. So, before undertaking a study in this field, the first thing to do is determine the kind of result that one aimes at. In this thesis, we take as starting point the work done by [13]: in that work, the aim is to estimate the radiation field of an instrument (specifically, a violin) in the context of an actual performance. This means that the instrument is played by a musician that performs a musical piece standing in front of an acquisition device. While it could seem like the most natural way of accomplish the task, this approach is in fact quite new in the literature. Capturing a real performance in order to estimate the radiation pattern is not an easy task, due to the fact that the musician is moving, and so is the instrument itself. All the existing works before [13] consider the instrument is in a fixed position and design the measurement system accordingly. A part of these works, such as [19], [4], [3],

[7], work in an "artificial" context: the instrument excitation is not provided by a musician, but through external loudpeakers pointed at the soundboard, mechanical bows or hammers that strike the instrument bidge. These approaches aim at estimating the radiation pattern with a high precision, in order to assess its acustical interaction with the surrounding environment only as a function of its phisical and mechanical characteristics. Also in [13] the aim is to study the interaction with the environment, but unlike the aforementioned methods, the instrument is regarded more as an acoustic entity that also interacts with the musician, rather than simply a sound source. Other works, such as [6], [7] and [14], follow the same idea: in these cases the presence of the musician actually influences the measurement, as his body absorbs or reflects a certain portion of the radiation of the instrument in a frequency-dependent fashion. Moreover, his arm provides (in the specific case through the bow) an excitation that is almost unreproducible artificially. However, as they again fix the instrument in space, the musician is not free to move: this fact greatly affects the naturalness of the perfomance, producing a result that is accurate in terms of measurements, but still not fully realistic. [13] overcomes this issue by exploiting a tracking system to determine the position in space of the instrument throughout the performance. By doing so, the resulting radiation pattern measurement takes into account all the aspects of a real performance. The tracking is performed though an infrared strucured light camera (*Microsoft Kinect*) and a gyroscope fixed on the instrument.

The acquisition of the soundfield produced by the instrument is accomplished by a grid of microphones (called *plenacoustic camera*), that captures the acoustic scene and analyzes it with the tools provided by *plenacoustic analysis*, a subject of research currently pursued at Politecnico di Milano. Plenacoustic analysis relies on a representation of the soundfield based on acoustic rays, which is mapped onto a series of *plenacoustic images*. These images, through the use of beamforming techniques, provide information about the energy coming from all directions towards the camera. The plenacoustic camera is therefore conceived as window on the acoustic scene, and, in our case, a point of view on the radiance pattern of the instrument. The idea in [13] is that by letting the musician move freely, different portions of the radiation pattern are presented to the plenacoustic camera; the more the musician moves, the wider is the observed pattern. This means that this aspect is not only beneficial under the points of view expressed so far, but is also a key requisite, since not all the pattern can be observed unless the musician moves enough. Freedom of movement might therefore turn into a restrictive constraint: the musician might have to move in an excessive way in order to allow the camera to pick the energy coming from the instrument at all angles, thus nullifying the very foundation of the system.

Our starting idea to improve that system is simple, and quite straightforward: the same system can be much more effective if two plenacoustic cameras are used simultaneously to observe the acoustic scene from two distinct points of view. In this way, two completely different portions of the pattern can be measured at a time, allowing the musician to move with far less contraints. We can observe how this immediately introduces a positive side effect: two points of view allow the system to acquire a double amount of data, rendering therefore the acquisition process much faster. The extension to a multiple-camera scenario of the system presented in is actually quite straightforward and only requires some minor modifications.

As a matter of fact, it only needs to integrate the information coming from the two cameras: the rest of the system (source tracking, pattern estimation) remains unvaried.

What we decided to explore, however, is a more specific aspect of this extension. We started by acknowledging that the system developed in [13] exploits only a very limited set of the tools provided by plenacoustic analysis. As we mentioned earlier, plenacoustic analysis relies on plenacoustic images: these images constitute a key tool, as they translate the complex information about the soundfield into a compact representation that, as well as being easy to inspect, enables the use of computer vision tools to perform advanced analysis of the acoustic scene. With some very easy-to-implement techniques, acoustic entities such as sources and reflector can be detected and localized on the acoustic scene. Therefore, in our case, besides being able to retrieve information about the energy coming from the instrument towards the camera, we could also be able to detect its presence and estimate its position in space, thus removing the need for an external tracking system. In the case of a single-camera based system, however, source localization proved to be not enough reliable due to the fact that the state of the art plenacoustic localization techniques, while providing a good estimation on the direction of the source, are not accurate in the estimation of the position when the source is far from the camera. Moreover, these techniques only work in two dimensions and are inapplicable to the context of a source free to move in three-dimensional space.

Our work aims at devising the multiple-camera plenacoustic tools necessary to build a fully plenacoustic radiation pattern estimation measurement system. We take as starting point the available pleancoustic imaging techniques first introduced in [12] (here presented in depth in Chapter 3), then investigates how the addition of a secondary camera affects those techniques. The multiple-camera paradigm is explored at different levels in Chapter 4. First, in Section 4.1, we see how two or more cameras interact in plenacoustic terms and how the pleancoustic images they generate can be combined into a single coherent view. We start from the two-dimensional acoustic rays-based representation of the soundfield (called *reduced ray space*) that was used so far to both display the plenacoustic images and perform analysis, and devise a new suitable way of managing it to fit the multiple-view model. We explain how this is done by passing to the full *projective ray space* for all the processing aspects, and how the reduced ray space can be considered only to display the plenacoustic images and perform the strictly computer vision-related part of the analysis.

The main implication of this change is presented in Section 4.2, where the single-camera source localization technique presented in Section 3.3 is extended to the multiple-camera case. The technique that we have devised retains the reduced ray-space representation, adequately modified to fit the multiple-view model, to display the plenacoustic image and obtain a first estimation of the source position, but takes into account the combined projective ray space for the actual estimation process. We explain the reasons for this change, presenting a series of issues deriving from the addition of more observation windows, as well as the course of action that we chose to address these issues. We also explain why the new method allows a more accurate source localization thanks to the different directional information obtained

through the multiple cameras composing the system.

Since the extension to the multiple-camera scenario requires that the position in space of the distributed cameras is known, in Section 4.3 we also introduce a method for calibrating the system (i.e. for estimating the relative position of the cameras) though a plenacoustic approach. At the end of the chapter, in Section 4.5, we finally extend the newly introduced concepts to a three-dimensional scenario: the three-dimensional source localization method takes a set of 2D localizations provided by the system and geometrically estimates the global 3D position. Once the multiple-camera paradigm has been explored, in Chapter 5 we show how the concepts apply to the radiation pattern estimation system, which we reformulate in order to take advantage of the newly devised pleanacoustic methods.

In Chapter 6 we first present a series of simulative and experimental results. The former, aim at testing the effectiveness of the multiple-camera plenacoustic methods devised in this work, the latter show the perfomance of our first implementation of the fully plenacoustic multiple-camera radiation pattern measurement system. This implementation exploits the newly introduced methods to measure the radiation pattern of an acoustic source in fixed position.

Before presenting our studies, in Chapter 2 we provide some theoretical background on acoustics and we devote a part of the thesis (specifically Chapter 3) to a detailed dissertation on the state of the art plenacoustic concepts that we use as a foundation for all our work.

Chapter 2

Background

In this first chapter we start by putting a mathematical foundation that will be useful for the rest of the thesis, and provide a synthetic dissertation on the concepts of farfield and nearfield radiation, directivity patterns and multipoles. We then proceed with an overview of some meaningful works related to the pattern radiation estimation problem that will serve as a basis for comparison with our work. We also give a first glimpse on plenacoustic analysis, a core aspect of our study that will be exhaustively covered in Chapters 3 and 4. Eventually, we provide a description of the concepts of projective geometry that we will use extensively in our work.

2.1 Foundamentals of Acoustics and Directivity Pattern

In this section we introduce the Rayleigh integral, which is a widely used model in the sudy of the radiation field, then we cover some basics on farfield radiation and multipoles.

2.1.1 Rayleigh Integral

Let us consider the *Kirchhoff-Helmholtz Integral*:

$$P(\mathbf{x},\omega) = \oint_{\partial S} \left(\frac{\partial}{\partial \mathbf{n}} G(\mathbf{x}|\xi,\omega) P(\xi,\omega) - G(\mathbf{x}|\xi,\omega) \frac{\partial}{\partial \mathbf{n}} P(\xi,\omega) \right) d\xi. \tag{2.1}$$

where $\mathbf{x} = (x, y, z)^T$ is a 3D point in spatial coordinates, ω the pulsation measured in rad/s and $P(\mathbf{x}, \omega)$ is the Fourier transform of the pressure $p(\mathbf{x}, t)$ [20]. $G(\mathbf{x}|\xi, \omega)$ is the *Green's frunction*, which for all sources at location $\xi \in S$ in free field is defined as

$$G(\mathbf{x}|\xi,\omega) = \frac{1}{4\pi} \frac{e^{-jk|\mathbf{x}-\xi|}}{|\mathbf{x}-\xi|}, \tag{2.2}$$

where $k = \omega/c$ is the wave number (with c speed of sound in the medium).

The Kirchhoff-Helmholtz Integral states that the sound pressure is completely determined within a volume free of sources, if sound and velocity are known on all points on its surface S. This means that any sound field can be reconstructed if the

sound pressure and the acoustic velocity are restored on all points of the surface of its volume (which is the underlying principle of *holophony*).

With some passages we can rewrite (2.1) as:

$$P(\mathbf{x},\omega) = \frac{1}{4\pi}\int_S \left(j\omega\rho_0 V_s(\omega)\frac{e^{-jkr}}{r} + P_S(\omega)\frac{\partial}{\partial \mathbf{n}}\frac{e^{-jkr}}{r}\right) dS \tag{2.3}$$

where ρ_0 is the density, r is the distance from the surface point to the observation point and V_S, P_S respectively the sound particle velocity and the pressure on S.

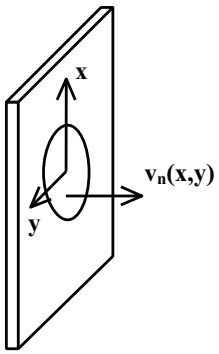


Figure 2.1. Baffled Piston Representation

Now, to approximate the behaviour of a source, we consider a *baffled piston* of area S mounted on an infinitely large hard wall in the x, y plane. The source generates plane waves and it is defined by the boundary condition that normal velocity is null ($v_n = 0$). So the contribution of the surface sound pression of the second addend of (2.1) is null, leading to

$$P(\mathbf{x},\omega) = \frac{j\omega\rho_0}{2\pi}\int_S \left(V_n(x,y,0,\omega)\frac{e^{-jkr}}{r}\right) dS. \tag{2.4}$$

The previous formula is referred to as the *Rayleigh Integral* and is widely used to study the radiation field of distributed acoustical source since it relates the velocity on the surface to the acoustic pressure radiated. Basically, it states that the radiation of any planar source is determined from the two-dimensional Fourier transform (on x, y) of its normal velocity distribution. Considering the spherical coordinates

$$\begin{aligned} x &= r\sin\theta\cos\phi, \\ y &= r\sin\theta\sin\phi, \\ x &= r\cos\theta, \end{aligned} \tag{2.5}$$

we can write (2.4) as

$$p(r,\theta,\phi) = \frac{j\omega\rho_0}{2\pi} D(\theta\phi). \tag{2.6}$$

where $D(\theta,\phi)$ is the *directivity function* defined as

$$D(\theta,\phi) = \frac{e^{-jkr}}{r}\mathcal{F}_x\mathcal{F}_y[v_n(x,y,0)] \tag{2.7}$$

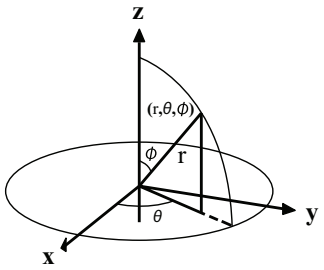


Figure 2.2. Spherical Coordinates

that, being phase and distance independent, is often used to plot the directivity characteristics of a source. The directivity function of acoustic pressure radiation is measured in Pascal-meters and specifies who the pressure is distributed as a function of θ, ϕ at a give distance r (i.e. on a sphere around the source), at a certain frequency ω. Once the Fourier transform of the velocity pattern is computed for a given source distribution on a plane, these equation provide the directivity for any frequency.

2.1.2 Farfield Radiation and Spherical Waves

Two main regions can be distinguished depending on the distance of the point of observation from the radiator. In the *nearfield* (or *Fresnel*) region, the radiator produces a more or less collimated beam which can be approximated as a perturbed plane wave. On the other hand, in the *farfield* (or *Fraunhofer*) region, the radiator produces a diverging spherical wave whose sound pressure magnitude decreases linearly with distance, and it can be approximated as a point source with a given radiation pattern.

The condition for a finite sized radiation to be considered a point source is that the distance r of the point of observation from the source is large with respect to both the emitted wavelength and the radius of the bounding sphere of the radiatior:

$$r >> \lambda \tag{2.8a}$$

$$r >> a. \tag{2.8b}$$

A point source emits ideal spherical waves: the spherical geometry provides a finite and compact expansion of wavefronts which allos us to gain an understanding of expanding waves. Furthermore, compact realistic vibrators are more closely modeled with sperical wave expansions that with planar expansions [20]. The acoustic wave equation given in spherical coordinates (2.5), for a homogeneous fluid with no viscosity, is

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial p}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 p}{\partial\phi^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0. \tag{2.9}$$

Its solution is given by separation of variables:

$$p(r,\theta,\phi,t) = R(r)\Theta(\theta)\Phi(\phi)T(t), \tag{2.10}$$

which leads to four ordinary differential equations:

$$\frac{d^2\Phi}{d\phi^2} + m^2\Phi = 0; \tag{2.11}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}sin\theta\frac{d\Theta}{d\theta} + \left[n(n+1) - \frac{m^2}{sin^2\theta}\right]\Theta = 0; \tag{2.12}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + k^2R - \frac{n(n+1)}{r^2}R = 0; \tag{2.13}$$

$$\frac{1}{c^2}\frac{d^2T}{dt^2} + k^2T = 0. \tag{2.14}$$

We can combine the angle functions (2.11) and (2.12) into a single function Y_n^m called *spherical harmonic*:

$$Y_n^m(\theta,\phi) \equiv \sqrt{\frac{(2n+1)}{d\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{jm\phi}, \tag{2.15}$$

where P is the Legendre polynomial. It can be demonstrated that any solution to (2.9) for traveling waves can be written as

$$p(r,\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{mn}h_n^{(1)}(kr)Y_n^m(\theta,\phi), \tag{2.16}$$

where A_{nm} are complex constats and $h_n^{(1)}$ is the spherical Hankel function of the first kind. Therefore, any arbitrary function on a sphere $g(\theta,\phi)$, like any farfield radiation pattern of a finite sized source, can be expressed with a combination of spherical harmonics

$$g(\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{mn}Y_n^m(\theta,\phi) \tag{2.17}$$

2.1.3 Multipoles

Radiation from bodies of finite extent can be characterized by sums of multipoles. We start by considering the simplest multipole, the *monopole*, which is an ideal point source. It is said to be *omnidirectional*, meaning that its associated radiation is independent of angle, like an ideal spherical harmonic of $m, n = 0$. Its pressure is given by

$$p = \frac{-i\rho_0 ck}{4\pi}Q_s\frac{e^{ikr}}{r}, \tag{2.18}$$

where Q_s is the volume flow. This formula is derived by considering a small oscillating spherical cavity in the farfield, and same assumptions can be made for any real source that can be approximated with a spherical cavity. We can extend what we have shown at the beginning of 2.1.2, by saying that the angular distribution of radiation for a finite sized source is governed by the relation that links the size of the radiator to $\lambda' = \lambda/2\pi$ (with λ the radiated wavelength). If the radiator is much smaller than λ', the details of its structure become unimportant: the radiator can

be approximated with a monopole, because parts of the surface that move outward are compensated by others which move inward at the same moment. On the other side, if the radiator is large, individual regions of the surface will radiate more or less independently, interfering with each other and producing "beams" that spread out with distance. The result is that radiation pattern gets more complex. For a sphere-like source, the transition from the "small radiator" to the "large radiator" regime occurs when its radius r is equal to the wavelength λ' emitted, so at a frequency $f = c/\lambda'$.
Sound fields other than omnidirectional can be obtained by superposing two or more monopoles. For simplicity, we can discuss the radiation field of a dipole, i.e. a pair of two identical point sources opposite in phase and separated by a short distance d. If we take $D_s = Q_s d$ (called dipole strength), from (2.4) we can obtain that the radiated pressure of a dipole is

$$p(r, \theta, \phi) = -\rho_0 c k^2 D_s \cos\theta \left(1 + \frac{i}{kr}\right) \frac{e^{ikr}}{4\pi r}. \tag{2.19}$$

Since the dipole is oriented along the z axis, the directivity pattern is independent of ϕ and depends only on $\cos\theta$ as shown in fig. 2.3, with a classic figure-8 pattern, like an ideal spherical harmonic of $m = 0$ and $n = 1$. In the same way, many monopoles can be arranged with different configurations in order to create multipoles with different radiation patterns, each defined by a different spherical harmonic.

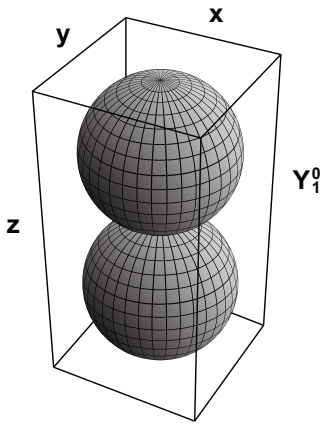


Figure 2.3. Dipole

While those that we have seen so far are ideal radiators, many compact acousic sources produce sound fields that can be represented in spherical coordinates in the same way as simple multipole sources. Also for very complex sources such as musical instruments, it is possibile to make deductions about farfield directivity based on some standard cases.

2.2 Related Work

We have already outlined in chapter 1 the work made in [13], which can be considered, for the present work, the state of the art. However, many approaches to

radiation pattern estimation have been devised in the literature over the years, each one based on different assumptions and aimed at different goals. The list of available examples is long and its in-depth analysis is out of the scope of our work. In this section we therefore present only a few meaningful examples of radiation pattern measurement systems, from both the hardware and computational point of views.

2.2.1 Measuring Musical Instruments Directivity Patterns With Scanning Techniques

The approach of [7] is to adopt a scanning techinque that uses a fixed reference microphone close to the source and a pressure microphone that moves across a planar surface at distance R from the source. The idea is that the transfer functions between the fixed and moving sensors give a time independent ratio which carries information about the directivity of the sound source. If we represent the sound source as a set of monopoles closely distributed in space, the pressure field can be described with

$$p(\theta, \phi, r, t) = \frac{\mathbf{A}}{r} D(\theta, \phi) e^{j(\omega t - kr)}, \tag{2.20}$$

where $\mathbf{A}$ represents the time-independent term that relates source characteristics (such as volume velocity, specific acoustic impedance and wave number) at given position θ, ϕ, r (azimuth, elevation, distance from source and measurement plane) and angular frequency ω. $D(\theta, \phi)$ is the directivity term that can be considered unitary for monopoles or equal to the first order Bessel function if the source behaves like a baffled piston. For musical instruments, though, it cannot be described using general analitic expressions: a specific measurement is needed to characterize complex sound sources. For the reference mic, they write the expression of pressure as:

$$p(\hat{\theta}, \hat{\phi}, \hat{r}, t) = p_{ref}(t) = \mathbf{B} e^{j(\omega t - k\hat{r})}. \tag{2.21}$$

Then, they map the values of the meausured pressures across the measurement plane onto a sphere of radius R:

$$\begin{aligned} p_{sph}(\theta, \phi, t) &= p(\theta, \phi, r, t) \left(\frac{R + \Delta r}{R} \right) e^{j\Delta r} \\ &= \frac{\mathbf{A}}{r} D(\theta, \phi) e^{j(\omega t - kr)} \left(\frac{R + \Delta r}{R} \right) e^{j\Delta r}, \end{aligned} \tag{2.22}$$

Δr being the euclidean distance between the corresponding sphere projection and the measurement position. At this point, to define time harmonic excitation, they relate (2.20) and (2.21) by computing the transfer function:

$$H_1 = \frac{\mathbb{F}\{p_{sph}(\theta, \phi, t)\}}{\mathbb{F}\{p_{ref}(t)\}} = D(\theta, \phi) \left(\frac{\mathbf{A}(\omega)}{\mathbf{B}(\omega)} \frac{e^{-jkR}}{R} \right). \tag{2.23}$$

Since H_1 is independent from the position of the moving sensor, it could be seen as a scaled version of the directivity term $D(\theta, \phi)$

$$H_1 = \gamma(\omega) D(\theta, \phi), \tag{2.24}$$

where $\gamma(\omega)$ is a scaling factor depending on the position of the reference sensor, the radius R and the source charateristics. Measuring the transfer function between fixed and projected signal leads to a time independent expression for calculating the directivity of any sound source.

2.2.2 Method for Measuring Violin Sound Radiation Based on Bowed Glissandi and Its Application to Sound Synthesis

The system implemented in [6], relies on a set of microphones arranged in a spherical-like fashion, with the source positioned at center of the sphere. The source is a violin in fixed position, played by a violinist. Calibration of the system is performed through lasers and an *OptiTrack* system of 6 cameras. To measure the directivity pattern of the violin, transfer functions between the microphones and a reference pickup placed at the bridge of the violin are computed. To do so, the violin Body Frequency Response (BFR) is obtained through signal deconvolution computed on a frame-by-frame basis. The violinist plays a very slow glissando on the G string. Both the reference signal $x(t)$ obtained with the bridge pickup (referenced to as the *excitation signal*) and the signals $y(t)$ obtained through microphones (*response signals*) are divided in frames that exhibit an almost-stationary content. After aligning in time the frames of $x(t)$ and $y(t)$ (the known reference-to-microphone delay is compensated), the signals are windowed and expressed in the spectral domain as $X(f,k), Y(f,k)$, with f, k being the frame index and frequency bin respectively. The frame-by-frame decovolution is computed as

$$\begin{aligned} |H_{BFR}(f,k)| &= \frac{|Y(f,k)|}{max(|X(f,k)|, b)}, \\ \varphi[H_{BFR}(f,k)] &= \varphi[Y(f,k)] - \varphi[Xs(f,k)] + 2\pi m, \end{aligned} \tag{2.25}$$

with b a threshold set to 10^{-5}. The magnitude is then averaged across all frames as

$$|H_{BFR}(k)| = \frac{\sum_{f=1}^{N} w_f(k) 20 \log_1 0 |H_{BFR}(f,k)|}{\sum_{f=1}^{N} w_f(k)}. \tag{2.26}$$

The weight $w_f(k)$ is the input signal energy at frame f and bin k. The ratio between BFRs shows the radiation ration at different directions.

2.2.3 Method for Measuring Acoustic Radiation Fields

In [18], the measurement of the radiation field is performed through a mechanical boom system that allows two microphones to move on two concentric spherical patterns centered at the source location. The aim is to measure the acoustic field that is here parameterized as an expansion in sperical waves of the acoustic pressure:

$$p(r,\theta,\phi) = \sum_{L,M} \Big(a_{L,M} h_L(kr) b_{L,M} h_L^H(kr) \Big) Y_{L,M}(\theta,\phi) \tag{2.27}$$

where $Y_{L,M}(\theta,\phi)$ is a sperical harmonic and $h_L(kr)$ a sperical Hankel function of the first kind, defined as

$$h_L(x) = (-x)^L \left(\frac{d}{dx} \right)^L \left(\frac{e^{jx}}{jx} \right). \tag{2.28}$$

The Hankel function represents an outgoing spherical wave, while its complex conjugate represents the relative incoming wave. The idea behind the method is that if we know the value of $p(R_1, \theta, \phi)$ at all points of a sphere of known radius R_1, the expansion of the function in spherical harmonics yelds values of $a_{L,M} h_L(kr) b_{L,M} h_L^H(kr)$ for every L, M. Repeating the measure on a second sphere of radius R_2 yelds another linear combination of the coefficients, which allows to compute their values. After completing the measurement with the microphones over various combinations of θ, ϕ, obtaining the complex amplitudes $F(\theta, \phi)$ of the microphones signals, the coefficient of the expansion can be computed as

$$C_{L,M} = \int_0^{\pi} \mathcal{P}(\theta) G(M; \theta) \sin\theta d\theta \tag{2.29}$$

where $\mathcal{P}$ is the normalized associated Legendre function and

$$G(M; \theta) = (2\pi)^{-\frac{1}{2}} \int_0^{2\pi} e^{-jM\phi} F(\theta, \phi) d\phi. \tag{2.30}$$

At this point, knowing the radii, frequency and speed of sound, the coefficients a, b can be computed from the expansion.

2.3 First Glimpses on Plenacoustic Analysis

It should be clear that no universal approach to the problem exists. All the techinques present in the literature differ in a variety of aspects. Mainly, they differ in the the mathematical models they exploit for their analysis, which, as a consequence, impose specific measurement systems. Our system relies on a plenacoustic representation of the soundfield. Plenacoustic analysis is a novel soundfield imaging method first proposed in [12], whose goal is to capture the acoustic radiance over a given aperture (Observation Window) facing the acoustic scene. The soundfield is captured by one or more microphone arrays (referred to as *plenacoustic cameras*) and is parameterized through a representation based on acoustic rays. Any ray is represented as a vector in a space called *ray space* of coordinates

$$\mathbf{l} = (l_1, l_2, l_3)^T \tag{2.31}$$

and is associated to an energy value computed through beamforming techniques. One of the key features of plenacoustic imaging is the compact visualization of this information: a reduced version of the ray space is visulized as a 2D image (called *plenacoustic image* or *soundfield image*) where acoustic primitives such as sources, receivers and reflectors are represented by geometric primitives such as lines and points. The plenacoustic image, in addition to providing a means to easily inspect this complex load of information, also allows us to use computer vision tools to perform a set of analyses on the soundfield in a very compact way. The problem of localizing a source or a reflector, as an example, reduces to the problem of estimating parameters of the geometric primitives in the image.

Clearly, plenacoustic analysis provides all the tools we need for our goal, since it allows us to measure the acoustic energy coming from all angles in proximity of the plenacoustic camera. We are then able to detect a sound source, estimate its

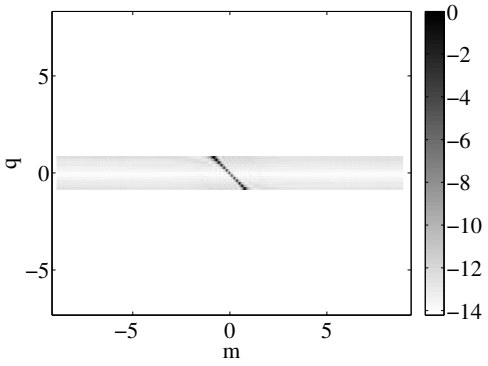


Figure 2.4. A Plenacoustic Image

position and evaluate the acoustic energy it radiates: we already have the building blocks to the system we want to implement.

While all these concepts will be covered in depth in the next chapters, this very brief introduction allowed us to introduce the representation of the soundfield based on acoustic rays. Since this aspect is the key to all our studies, we provide also a review of projective geometry, which is essential for managing acoustic rays.

2.4 Projective Geometry

To manage the soundfield representation based on acoustic rays, we need projective geometry. To introduce projective geometry, we start from the familiar Euclidean geometry. Euclidean geometry is troublesome in one major aspect: we need to keep making an exception to reason about some of the basic concepts, such as intersection of lines. Two lines almost always meet in a point, except if they are parallel. This exception is commonly worked around by stating that such parallel lines meet at infinity. However this statement conflicts with the dictum that infinity does not exist. We therefore enhance the Euclidean plane by adding those points at infinity where parallel lines meet, and resolve the difficulty with infinity by calling them *ideal points*. Under this perspective, the projective space is just an extension of the Euclidean space in which two lines alway meet in a point, which may be at infinity.

2.4.1 Homogeneity, Points and Lines

Part of this extension is the adoption of *homogeneous coordinates*. A point in the projective space is represented by a *equivalence classes* of coordinate triples, where the two triples $\mathbf{p} = (x, y, 1)^T$, $k\mathbf{p} = (kx, ky, k)^T$ represent the same point for any $k \neq 0$. Given (kx, ky, k) we can get its inhomogeneous coordinates back by dividing by k and removing the last coordinate. More generally:

$$\mathbf{p} = (x_1, x_2, x_3)^T \iff \boldsymbol{p} = \left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right)^T. \tag{2.32}$$

Defining points at infinity should now be straightforward: a point a infinity is any point whose third homgeneous coordinate is 0, which corresponds to a point $\boldsymbol{p} = (x_1/0, x_2/0)^T$. The points at infinity in the two-dimensional projective space form a line called *line at infinity*. However all the considerations made so far are not bound to two dimensions: by representing points as homogeneous vectors, any Euclidean space $\mathbb{R}^n$ can be extended to a projective space $\mathbb{P}^n$. Note that throughout this thesis we adopt the convention of representing points in $\mathbb{R}^2$ (i.e. in *inhomogeneous coordinates*) with an italic-bold face, while points in $\mathbb{P}^2$ with a bold face as done in equation (2.32).
The same concept can be applied to lines. Equations $l_1x + l_2y + l_3 = 0$ and $kl_1x + kl_2y + kl_3 = 0$ (with $k \neq 0$) represent the same line on a plane. Thus any line can be identified by the vectors $\mathbf{l} = (l_1, l_2, l_3)^T$ and $k\mathbf{l} = k(l_1, l_2, l_3)^T$. Again, two such vectors related by an overall scaling are considered as beign equivalent. We can therefore state that the set of equivalence classes of vectors in $\mathbb{R}^3 - (0, 0, 0)^T$ forms the projective space $\mathbb{P}^2$ (the point $(0, 0, 0)^T$ is excluded since it does not represent any line nor point).
Adopting homogeneous vector representations for lines and points allows us to perform some meaningful operations in a very compact way. As an example, we can state that a point $\mathbf{p}$ lies on a line $\mathbf{l}$ if and only if

$$\mathbf{p}^T\mathbf{l} = 0, \tag{2.33}$$

i.e. if and only if their inner (scalar) product is equal to zero. Also the intersection between lines is quite straighforward: two lines $\mathbf{l}$ and $\mathbf{l}'$ intersect in the point

$$\mathbf{p} = \mathbf{l} \times \mathbf{l}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ l_1 & l_2 & l_3 \\ l'_1 & l'_2 & l'_3 \end{vmatrix} \tag{2.34}$$

where the $\times$ operand stands for vector (cross) product.

2.4.2 Projectivies and Isometries

Projectivities (also called *projective tranformations* or *homographies*) are mappings of points in $\mathbb{P}^2$ to points in $\mathbb{P}^2$ that preserve lines. It is possible to prove that any linear transformation on homogeneous 3-vectors represented by a non-singular 3×3 matrix is a projectivity. Let $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ lie on a line $\mathbf{l}$, i.e. $\mathbf{l}^T\mathbf{p}_i = 0$ for $i = 1, 2, 3$. If $\mathbf{H}$ is a non-singular 3×3 matrix, we can observe that $\mathbf{l}^T\mathbf{H}^{-1}\mathbf{H}\mathbf{p}_i = 0$. Thus the points $\mathbf{H}\mathbf{p}_i$ all lie on the line $\mathbf{H}^{-T}\mathbf{l}$, and collinearity is preserved by the transformation.
A generic projectivity can be therefore written as

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{2.35}$$

or more compactly

$$\mathbf{p}' = \mathbf{H}\mathbf{p}. \tag{2.36}$$

As for lines, we have just seen that if points $\mathbf{x}_i$ lie on a line $\mathbf{l}$, then the transformed point lie on the line $\mathbf{l}' = \mathbf{H}^{-T}\mathbf{l}$. Thus under the point transformation $\mathbf{p}' = \mathbf{H}\mathbf{p}$, a line $\mathbf{l}$ transforms as

$$\mathbf{l}' = \mathbf{H}^{-T}\mathbf{l}. \tag{2.37}$$

In our thesis we limit our focus on projectivites to the special case of *isometries*. Isometries are defined as transformations of the plane $\mathbb{R}^2$ that preserve Euclidean distance and are represented as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & \Delta x \\ \sin\theta & \cos\theta & \Delta y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.38}$$

or more concisely in block form as

$$\mathbf{p}' = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{p} \tag{2.39}$$

where $\mathbf{R}$ is a 2×2 orthogonal rotation matrix and $\mathbf{t}$ a translation 2-vector. Being a composition of a rotation and a translation and orientation-preserving, isometries are defined as Euclidean transformations.

2.5 Conclusions

This chapter served as a first overview of the various aspects related to our work. Even though the underlying mathematical model is very complex and not easy to manage, the choice of plenacoustic analysis allows us to address the problem of radiation pattern estimation in an incredibly compact way. However, only a very limited set of the tools provided by plenacoustic analysis were exploited in [13]: our main goal in this thesis is to dig further in the possibilities provided by this kind of analysis, see how the plenacoustic tools can be used in a more effective way in the context of the radiation pattern estimation problem. We start from the idea that using more than one plenacoustic camera enables a faster and easier acquisition process with respect to [13], then we explore how this choice impacts on the available plenacoustic techniques. In the next chapter we provide an overview of plenacoustic analysis as it was proposed in [12], then we proceed with the description of the core aspect of our studies: multiple-camera plenacoustic imaging.

Chapter 3

Single-Camera Plenacoustic Imaging

In section 2.3 we provided a first introduction on plenacoustic analysis and explained why it is suitable for addressing the radiation pattern estimation problem. In [13], plenacoustic analysis is exploited only for evaluating the energy radiated from the source, while in our work we intend to use it in wider scope. In this chapter we present the state of the art of plenacoustic analysis as it was introduced in [13], focusing on the aspects related to our ploblem. This will set a foundation for the the next chapter, where we will present the improvement and modifications that we have apported to these concepts and techniques in order to accomodate the multiple-camera scenario that we intend to adopt.
First we introduce the plenacoustic function, whose measurement through an array of microphones allows us to perform a detailed analysis of the acoustic scene. Then we explain the key concepts of ray space and plenacoustic image, as well at the techinque for single-camera source localization that we will later expand to the multiple-camera case.

3.1 Plenacoustic Function

The plenacoustic function can be thought of as a parameterization of the sound field, and it is a function that describes the acoustic radiance in every direction through every point in space. This means that, in the case of a 2D geometric domain, it can be written as a function $f(x, y, \phi, \omega, t)$ of position (x, y); direction ϕ; frequency ω and time t. In order to measure the plenacoustic function in a single point we can center a microphone array (an *acoustic camera*) in that location and estimate the acoustic radiance along all look directions through beamforming. Extending this measurement to a larger observation window (e.g. a segment in 2D geometry) would mean performing the same acoustic radiance estimation process on a whole set of points. To do so, a new kind of device is needed: we refer to this extension of the acoustic camera as the *plenacoustic camera* or *soundfield camera*. As a first step, however, we then need a simple way to parameterize and represent in a compact way the data that this device would capture, in order to be able to analyze the substantially larger information gathered on the global acoustic scene. In order to

get a more compact representation we drop the dependecy on time (t) and frequency (ω). If the hypotheses of validity of geometrical acoustics hold, we can adopt a representation based on *acoustic rays*.

3.1.1 The Ray Space

An acoustic ray is defined as an oriented line that identifies a planar acoustic wavefront component and is inherently perpendicular to it. This means that it can be graphically represented simply by means of a line $l_1x + l_2y + l_3 = 0$ of parameters $\mathbf{l} = (l_1, l_2, l_3)^T$. We here define the *ray space* $\mathcal{P}$ as the projective space $\mathbb{P}^2$ of points of coordinates $\mathbf{l} = (l_1, l_2, l_3)^T$ that correspond to lines with same parameters in the geometric space. Recall that, as stated in 2.4, for any $k \neq 0$, the two vectors $\mathbf{l}$ and $k\mathbf{l}$ represent the same line, i.e. the same ray. However, since we need to take into account also the orientation of rays, we adopt the convention of representing rays going towards the observation window from the positive half-space with $k > 0$ and rays going in the opposite direction with $k < 0$.

Rather than visualizing the whole 3D ray space, [12] resort to the *reduced ray space*, which is the 2D space obtained by cutting the ray space with a plane. As an example, if we choose to cut with the $l_2 = 1$ plane, every point in the reduced ray space is described by the coordinates $(m, q) = (-l_1/l_2, -l_3/l_2)$ which, of course, corresponds to representing lines in the geometric space in their slope–intercept form $y = mx + q$. Note that the information about orientation is lost in this simplification, as aray travelling on the two directions of the same line are represented with the same parameters. We deal with this issue by only taking into account those rays which are directed to the observation window (OW) from the positive half-space, i.e. those with $k > 0$.

3.1.2 Parameterization of a Source

Representing acoustic primitives such as sound sources and receivers in the ray space means defining the set of rays that are originated or received at their location. As an example, let us analyze a sound source, which is of capital interest in our case. A source can be represented in the geometric space by a point A of inhomogeneous coordinates $\boldsymbol{p}_A = (x_A, y_A)^T$ and homogeneous coordinates $\mathbf{p}_A = (x_A, y_A, 1)^T$. Equivalently, it can be thought as the set of all lines passing through it. The region of the ray space describing the parameters of such lines is called the *dual* $\mathcal{I}_{\boldsymbol{p}_A}$ of the source and is defined as

$$\mathcal{I}_{\boldsymbol{p}_A} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}_A^T \mathbf{l} = 0\}, \tag{3.1}$$

which corresponds to a plane passing through the origin and with parameters specified by the vector $\mathbf{p}_A$. This plane divides $\mathcal{P}$ in two half-spaces $\mathcal{P}^+_{\boldsymbol{p}_A}$ and $\mathcal{P}^-_{\boldsymbol{p}_A}$ defined as

$$\mathcal{P}^+_{\boldsymbol{p}_A} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}_A^T \mathbf{l} > 0\} \tag{3.2a}$$

$$\mathcal{P}^-_{\boldsymbol{p}_A} = \{\mathbf{l} \in P | \mathbf{p}_A^T \mathbf{l} < 0\}. \tag{3.2b}$$

$P^+_{\boldsymbol{p}_A}$ corresponds to all the rays that, along their travel, leave $\boldsymbol{p}_A$ on the left, whereas $P^-_{\boldsymbol{p}_A}$ corresponds to those that leave $\boldsymbol{p}_A$ on the right. We can see that in the reduced ray space, the plane $\mathcal{I}_{\boldsymbol{p}_A}$, being cut with another plane, becomes a line.

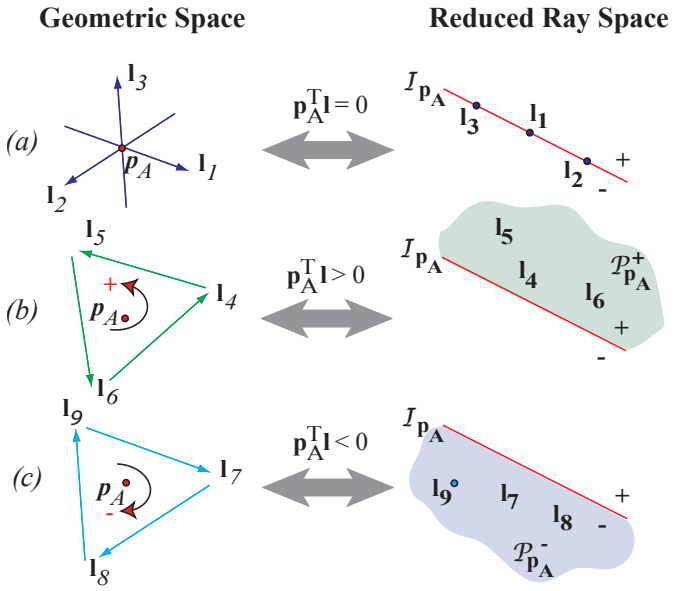


Figure 3.1. Source and Rays in Geometrical and Reduced Ray Space

3.1.3 Visibility Region and Region Of Interest

The OW can be seen in the geometric space as a segment $\boldsymbol{p}_A\boldsymbol{p}_B$, whose endpoints $\boldsymbol{p}_A$, $\boldsymbol{p}_B$ have as duals in the ray space the planes $\mathcal{I}_{\boldsymbol{p}_A}$, $\mathcal{I}_{\boldsymbol{p}_B}$. The rays crossing the OW can be of two types: 1) those that have $\boldsymbol{p}_A$ on the left and $\boldsymbol{p}_B$ on the right and 2) those that have $\boldsymbol{p}_A$ on the right and $\boldsymbol{p}_B$ on the left. Recalling the definition of $\mathcal{P}^+_{\boldsymbol{p}}$ and $\mathcal{P}^-_{\boldsymbol{p}}$ for a generic point $\boldsymbol{p}$, it is easy to verify that images of the rays of the two type are respectively:

$$\mathcal{I}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P}^-_{\boldsymbol{p}_B}\mathcal{P}^+_{\boldsymbol{p}_A} \tag{3.3a}$$

$$\mathcal{I}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P}^-_{\boldsymbol{p}_A}\mathcal{P}^+_{\boldsymbol{p}_B} \tag{3.3b}$$

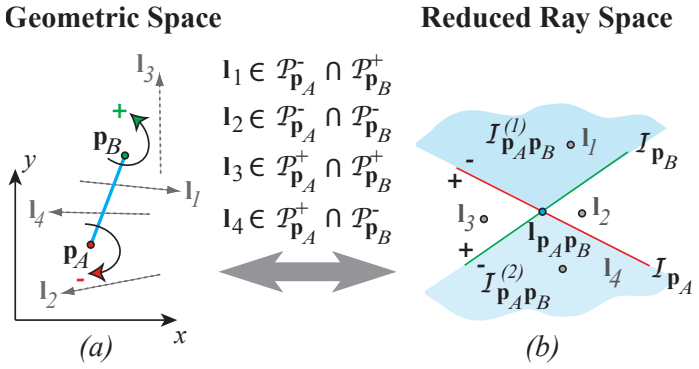


Figure 3.2. Visibility Region of an OW

The two wedge-shaped regions correspond therefore to the rays directed towards the two sides of the OW. We define with *visibility region* $\mathcal{V}$ the wedge-shaped region corresponing to the set of rays that cross the OW from the sensitive side of the aperture. If we want to assess what portion of the radiance produced by an acoustic primitive (e.g. a sound source) can be picked up by the OW, we only need to find which of the rays coming from it fall into $\mathcal{V}$. The set of these rays is referred to as the *Region Of Interest* (ROI) of a source $\boldsymbol{p}$, and is defined as:

$$\mathcal{R}_{\boldsymbol{p}} = \mathcal{V} \cap \mathcal{I}_{\boldsymbol{p}}. \tag{3.4}$$

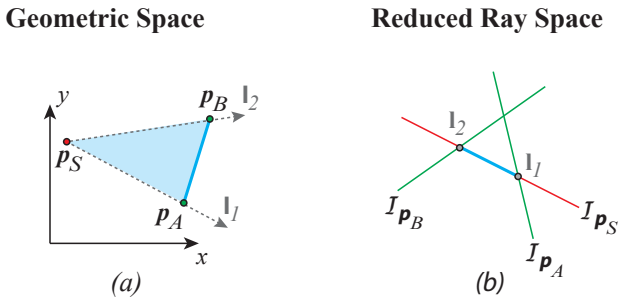


Figure 3.3. Region of Interest of a Source

3.2 Measurement of the Plenacoustic Function

In this section we introduce the concept of *plenacoustic image* (or *ray-space image*) as the ray-space representation of the sound field and we explain how this image is computed. As stated in 3.1.1, to visualize of the ray space it is necessary to cut it with a prescribed plane in order to lose one dimension and be able to report the information on a 2D image. We here adopt the same notation that is used in [12]: the cutting plane is $l_2 = 1$ so that the reduced ray space corresponds to the (m, q) space. The OW is considered to lie on the y axis between q_0 and $-q_0$, having therefore a strip-shaped visibility region.

3.2.1 The Plenacoustic Image

The plenacoustic function $f(x, y, \theta)$ can be mapped onto the ray space by imposing $x = 0$ (the OW is on the y axis), $\theta = \arctan(m)$, $-\pi/2 \leq \theta \leq \pi/2$ and $y = q$, leading to $p(m, q) = f(0, q, \arctan(m))$. For our purpose we discard the phase information and work with power images such as $P(m, q) = |p(m, q)|^2$.
Consider a source at $\mathbf{p}$ and its ROI $\mathcal{R}_{\boldsymbol{p}}$ in the reduced ray space: as the points of the segment correspond to the only rays that illuminate the OW, these are the only points where the ideal ray-space image takes on non-zero values. The pleanacoustic function in those points can be determined using the radiance beampattern $b_{\boldsymbol{p}}(\theta)$ of the source, which is the distribution of acoustic radiance produced by the source, as a function of the angle $\theta = arctan(m)$. The invariance of the acoustic radiance

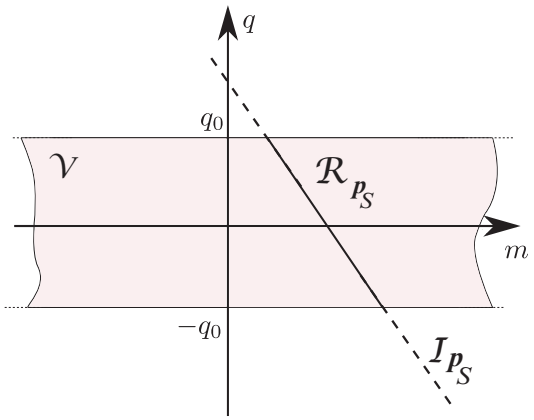


Figure 3.4. Reduced (m, q) space

along the ray allows us to write

$$p_{\boldsymbol{p}}(m,q) = \begin{cases} b_{\boldsymbol{p}}(\arctan(m)) & (m,q) \in \mathcal{R}_{\boldsymbol{p}} \\ 0 & \text{elsewhere} \end{cases} \tag{3.5}$$

In the case of multiple sources the complex contributions of each of them sum up to form the plenacoustic image. The details of the effects of reflection, though, are not very relevant to our problem, so we refer to [12] and [15] for this matter. The plenacoustic image that we just described is obviously a simplification of reality since both the scene is idealized (no diffraction or diffusion phenomena) and the camera is idealized (no issues of limited resolution or aliasing phenomena). The former aspect can be neglected, as in [12] it was assessed from preliminary measurements that the diffraction and diffusion generates features in the ray-space image whose magnitude is small with respect to direct path radiation or first-order reflections. The latter, on the contrary, needs further investigation.

3.2.2 The Real Plenacoustic Camera

Following the idea presented in section 3.1, a real plenacoustic camera can be thought of as an array of acoustic cameras placed on a grid that samples the OW. If the acoustic scene is static and the source signal stationary, we could employ a single acoustic camera and slide it along the OW. However, since it is safe to assume that in a real situation those two conditions are unlikely to be satisfied, we can exploit a single spatially extended microphone array and resort to a single-shot acquisition of the acoustic scene. The configuration best suited to our studies is a Uniform Linear Array (ULA). Let us consider an array of M microphones placed on the y axis between $y = q_0$ and $y = -q_0$. The wth microphone is in

$$\boldsymbol{m}_w = \left(0, \quad \frac{q_0 - 2q_0(w-1)}{M-1}\right)^T, \quad w = 1, \ldots, W. \tag{3.6}$$

In a sub-array of J (odd) microphones centered at $\boldsymbol{m}_w$, the sensors are located at $\mathbf{m}_j$ and acquire signals $s_j(t)$, where the index j is

$$j = w - \frac{J-1}{2}, \ldots, w + \frac{J-1}{2}. \tag{3.7}$$

We process $s_j(t)$ with a filter bank to obtain $s_j(\omega_k)$, ω_k being the the central frequency of the kth suband ($k = 1, \ldots, K$). The obtained signals are used to compute the psuedospectra $\mathbf{h}_w(\theta, \omega_k)$ through a Minimum Variance Distortionless Response (MVDR) beamformer for each sub-array and for each frequency ω_k. Depending on the needs, the wideband version of the pseudospectrum can be obtained as

$$H_w(\theta) = \prod_{k-1}^{K} \mathbf{h}_w(\theta, \omega_k), \quad w = \frac{J+1}{2}, \ldots, W - \frac{J+1}{2}, \tag{3.8}$$

w being the index of the sub-array and θ the direction of arrival (on the derivation of $\mathbf{h}(\theta, \omega_k)$ we invite the reader to refer to appendix A). Once computed, the single-band or wide-band pseudospectra must be mapped onto the ray space. We recall that the pseudospectrum $H_w(\theta)$ measures the power distribution of rays passing through the location $\boldsymbol{m}_w$ of the wth microphone. An acoustic ray passing through $\boldsymbol{m}_w$ and with angle θ has parameters

$$m = \tan(\theta) \qquad , \qquad q_w = q_0 - 2q_0 \frac{w-1}{W-1} \tag{3.9}$$

therefore we can write

$$\begin{aligned} P(m, q_w, \omega_k) &= \mathbf{h}_w(\arctan(m)), \\ P(m, q_w) &= H_w(\arctan(m)) \end{aligned} \tag{3.10}$$

which are respectively the wide-band and single-band (centered at ω_k) plenacoustic images. The former work directly on the pseudospectra and is useful when measuring the energy coming from a source at known θ, the latter performs a weighing on the sub-bands and, being more robust, is mainly used for source localization (we cover this topic in the next Section).

The real plenacoustic images that we obtain differ from the ones we would obtain with the ideal pleacoustic camera introduced in section 3.1 for two reasons: it is sampled along q (due to limited number of sub-arrays) and it is blurred along m (due to the limited number of sensors in each sub-array). In Figure 3.5 we show how a real plenacoustic image looks like. The image exhibits a ridge in the same location of $\mathcal{I}_{\boldsymbol{p}}$, which is the dual of the source $\boldsymbol{p}$. The ridge is, in fact, a blurred version of the visible portion of $\mathcal{I}_{\boldsymbol{p}}$ and the magnitude of the blurring varies with both q and m. This is due to the fact that the ULA does not exhibit a uniform resolution over θ: the farther the point from the source, the larger the incidence angle, the greater the blurring. Another visible anomaly is the large bright area in the lower left corner of the image: this is caused by angular aliasing, which is a consequence of the fact that the signal has frequency content that goes beyond the spatial Nyquist frequency (again, please refer to appendix A for further details on the spatial signal processing concepts).

3.3 Single Camera-Based Localization

One of the key analytic tools provided by plenacoustic analysis is source localization. To our purposes, being able to estimate the source position is essential, since

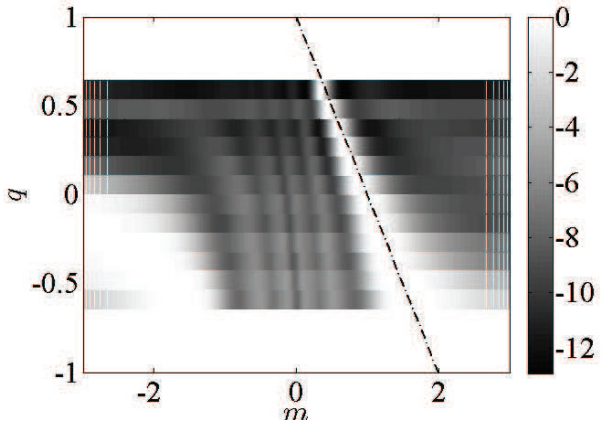


Figure 3.5. Real Plenacoustic Image

it allows us to drop the need for external devices to perfom the task exploiting the same set of tools that we use for radiation pattern estimation. The technique here presented is referred to a single-camera scenario, as it was introduced by [12]. In the next chapter we will explain how this techinque is expandable to a multiple-camera scenario.

Let us consider again the real pleanacoustic image shown in figure 3.5. It should now be clear that if we didn't know a priori the source position we could estimate it by finding the parameters of the line on which the ridge appears to lie, which is, as we now know, the reduced representation of dual plane $\mathcal{I}_{\boldsymbol{p}}$ of the source itself. While by direct inspection it's quite easy to detect this line, the accurate estimation of its parameters is not so immediate. A first approach is to find a global maximum of the acquired image and threshold all the values proportionally to this maximum in order to find peaks: the rays associated to these peaks should be the ones that are originated at the source location. Through a linear regression on the coordinates of the rays we could find an approximation of the lines parameters. However it is clear that this approach would not be effective, as an example, in the case of figure **??**: peaks would be detected also on the aliasing area, where the magnitude of the rays power is visibly comparable to the magnitude of the rays we are actually interested in. This, along with the possible presence of peaks generated by noise, makes the simple tresholding approach inapplicable.

The solution to this issue is found in the Hough Transform, a technique that recognizes special configurations of points in an image representable as a matrix exploiting their parametric representation. The Hough Transform estimates the parameters $\hat{x}, \hat{y}$ of $\mathcal{I}_{\boldsymbol{p}}$, thus determining the set I of indices that identify the rows of the ray-space image where the source is present (visible). We then determine the set of maxima (one per row) on the image:

$$\mathcal{L} = \left\{ (m_i, q_i) : \frac{|m_i \bar{x} - \bar{y} + q_i|}{\sqrt{1 + m^2}} < \epsilon, \quad i \in I \right\} \tag{3.11}$$

where i is the index of the sub-array (row of the ray-space image) and ϵ a given threshold. Let us consider a source in $\boldsymbol{p} = (x, y)^T$. We know from 3.1.2 that the set of rays departing from it is described by the equation $\mathcal{I}_{\boldsymbol{p}} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}^T \mathbf{l} = 0\}$. If

we consider the reduced ray space (m, q), the same set of rays is described by the equation $mx - y + q = 0$, which can be rewritten as

$$\mathbf{m}^T \boldsymbol{p} = -q, \quad \mathbf{m} = [m, -1]^T. \tag{3.12}$$

Then, the system of equations

$$\begin{cases} \mathbf{m}_{i_1} \boldsymbol{p} = -q_{i_1} \\ \vdots \\ \mathbf{m}_{i_N} \boldsymbol{p} = -q_{i_N} \end{cases} \tag{3.13}$$

is defined, where $i_1, \ldots, i_N$ are the indices in I. The system can be rewritten in matrix form as $\mathbf{M}\boldsymbol{p} = \mathbf{q}$. At this point, the estimate $\boldsymbol{p}$ can be found through a linear regression

$$\hat{\boldsymbol{p}} = (\hat{x}, \hat{y})^T = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{q}. \tag{3.14}$$

Note that this approach indirectly estimates the parameters of the plane $\mathcal{I}_{\boldsymbol{p}}$ relying on its reduced representation.

3.4 Conclusions

We have just introduced a series of concepts that constitute the very foundation of all the studies we have made in this thesis. Plenacoustic analysis is the cardinal element of both the radiation pattern measurement and the source localization problems. The parameterization of the plenacoustic function introduced in 3.1 is an essential starting point for addressing the problems at hand: the ray space-based representation enables an easier and more immediate analysis of the acoustic scene, and provides a solid basis for an advanced processing of the data acquired by the array. Furthermore, the source localization technique presented in 3.3 is a solid starting point for devising a system capable of localizing the source with the same data used for the radiation pattern measurement. In the next chapter we delve into the details of our studies. We start from the problem of finding a suitable representation of rays and we apply this representation to the problem of acquiring a plenacoustic image from multiple views. We then discuss the problem of localizing sources from such images and, finally, introduce the problem of localizing sources in the three-dimensional space from multiple 2D localization.

Chapter 4

Multiple-Camera Plenacoustic Imaging: Localization and Autocalibration

As already explained in the introduction to this thesis, the ability to simultaneously observe the source from different points of view is beneficial to the radiation pattern estimation problem. While the details of this aspect will be explained later in Chapter 5, in this chapter we explore what has been the actual focus of our study: how the introduction of a secondary camera affects the plenacoustic imaging techniques introduced in [12]. After a discussion about the modifications that we need to bring in order to acquire plenacoustic images with multiple cameras, we focus on the problem of source localization from multi-view images. In particular, we see how to manage the multiple OWs both in the geometric space and in the projective ray space, knowing their position and pose relative to a reference frame. Also, we explain how the information about their position can be retrieved through an autocalibration method that we have developed by exploiting the newly acquired multi-view tools. Being the presented methods based on plenacoustic techniques that work in two dimensions, the whole study was made under the same conditions and is therefore not applicable to 3D localization problems. To overcome this issue and be able to actually make use of these techniques in the context of a radiation pattern measurement system, we devised a method that estimates the three-dimensional position of the source based on the two-dimensional estimates provided by single arrays individually.

4.1 From Single-Camera to Multiple-Camera Plenacoustic Imaging

When dealing with multimple cameras we are as a first thing dealing with multiple reference frames, one attached to each camera. This means that, besides the global reference frame (x, y) we need to introduce the local reference frames $(x^{(i)}, y^{(i)}), i = 1, \ldots, C$, with C being the total number of cameras present in the acoustic scene. The superscript (i) is used from here on to specifiy that we are working in the ith camera reference frame. All the concepts introduced in chapter

3 are clearly still valid and can be applied to every single camera in our system. We only need to consider them as a series of different OWs, each with a local projective ray space $\mathcal{P}^{(i)}$ and its own local visibility region $\mathcal{V}^{(i)}$. In this section we explain how we have extended the concepts presented for the single-camera case to a multiple-camera case. First of all we illustrate how the local information of a camera can be combined with the information derived by all the other cameras composing the system. Secondly, we see how this new set of information is visualized on the plenacoustic image. At the end of the section we explain how the acquisition and combination of the plenacoustic images is performed.

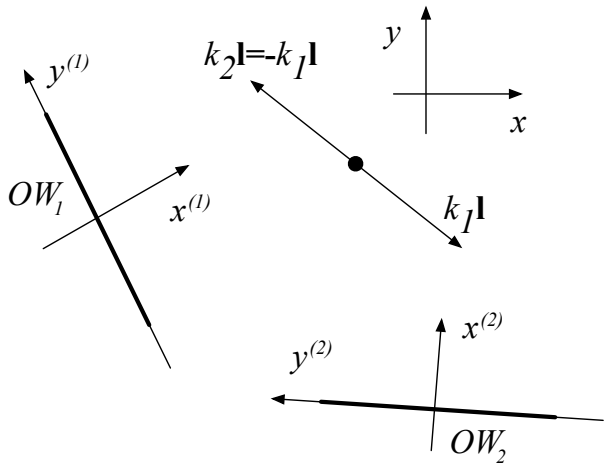


Figure 4.1. Global and Local Reference Frames

4.1.1 Fusion of Visibility Regions of Different Observation Windows

Every camera observes the acoustic scene from its point of view. It is clear that, to work with the information provided by different cameras, we need to combine these points of view into a single coherent view. Specifically, we need to bring all the local visibility regions into a global reference frame in order to have a shared point of view that we are able to analyze in its entirety. This is done by transforming each ray in the local visibility region through an homography. Referreing to what was explained in 2.4.2, a generic ray $\mathbf{l}^{(i)} = (l_1^{(i)}, l_2^{(i)}, l_3^{(i)})^T$ in the ith OW local reference frame can be related to the ray $\mathbf{l}_i = (l_{1_i}, l_{2_i}, l_{3_i})^T$ in the global reference frame by:

$$\mathbf{l}_i = \mathbf{H}^{(i)^{-T}} \mathbf{l}^{(i)}, \tag{4.1}$$

where

$$\mathbf{H}^{(i)} = \begin{pmatrix} \mathbf{R}^{(i)} & \mathbf{t}^{(i)} \\ \mathbf{0} & 1 \end{pmatrix}. \tag{4.2}$$

$\mathbf{R}^{(i)}$ and $\mathbf{t}^{(i)}$ are respectively the two dimensional rotation matrix and the translation vector that lead from the ith local to the global reference frame. Notice that in equation (4.1) we consider the inverse-transpose of the matrix $\mathbf{H}^{(i)}$: the reason to

this is explained (and proven) in section 2.4.2.

Now, if $\mathcal{V}^{(i)}$ is the visiblity region of the ith OW in the local reference frame and $\mathcal{V}_i$ is the same visibility region considered in the global reference frame, we can state that the global visibily region is the union of all $\mathcal{V}_i$

$$\mathcal{V} = \bigcup_{i=1}^{N} \mathcal{V}_i \tag{4.3}$$

In the global reference frame, the ROI of a source $\boldsymbol{p}$ is therefore defined as

$$\mathcal{R}_{\boldsymbol{p}} = \bigcup_{i=1}^{N} \mathcal{R}_{\boldsymbol{p},i} = \bigcup_{i=1}^{N} V_i \cap \mathcal{I}_{\boldsymbol{p}}, \tag{4.4}$$

where $\mathcal{I}_{\boldsymbol{p}}$ is the dual of the source $\boldsymbol{p}$ in the gloal reference frame and $\mathcal{R}_{\boldsymbol{p},i}$ the ROI of the source over the ith OW visibility region. Now, in order to visualize the global plenacoustic image, referred to as the *joint plenacoustic image*, we cut the global projective ray space with a prescribed plane, as explained in 3.2.

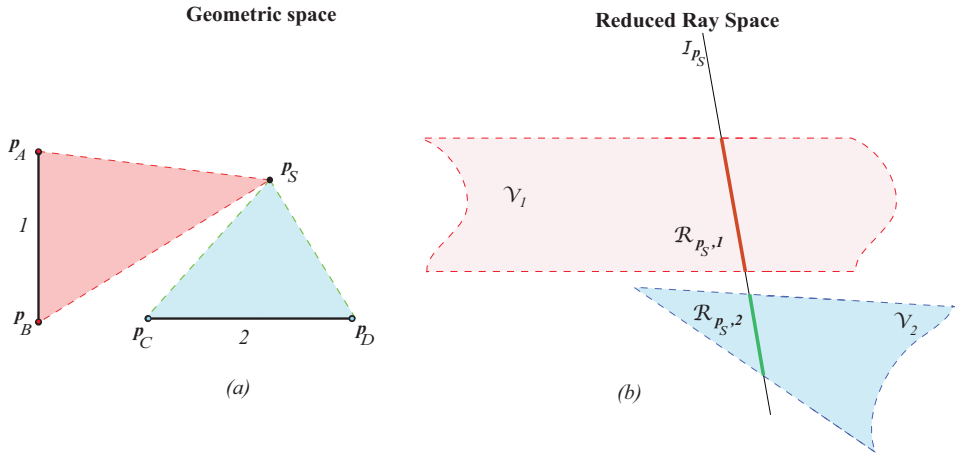


Figure 4.2. Acoustic Scene

4.1.2 Joint Plenacoustic Image

We saw in section 3.2.1 that the plenacoustic image is the mapping of the plenacoustic function onto the ray space. In the single-camera case, as done in [12], the mapping is performed directly onto the specific reduced ray space (m, q). However, adopting the slope-intercept representation $y = mx + q$ of rays prevents the ability of representing rays that are parallel to the y axis. In the single-camera case that is not an issue: since the global and local reference frame of the OW coincide, the rays parallel to the y axis are also parallel to the OW and cannot therefore be picked up by the camera, thus removing the need of representing them. The extension to the multiple-camera case on the other hand requires that those rays are representable as well, since they might be picked up by some camera in

arbitrary position in the global reference frame. We therefore need to explain how the mapping of the plenacoustic function is made onto the projective ray space. Consider a ray $\mathbf{l}$ passing through $\boldsymbol{p} = (x, y)^T$ with direction θ. Its parameters are

$$\begin{aligned} l_1 &= k\sin(\theta), \\ l_2 &= -k\cos(\theta), \\ l_3 &= k\big(y\cos(\theta) - x\sin(\theta)\big), \end{aligned} \tag{4.5}$$

with $k > 0$. The plenacoustic image $p(\mathbf{l})$, again defined as the mapping of the plenacoustic function onto the ray space, is

$$p(\mathbf{l}) = \begin{cases} f\Big(x, -\frac{l_1 x + l_3}{l_2}, -\mathrm{sgn}\big(\sin(\theta)\big)\arccos(l_2)\Big), & l_2 \neq 0 \\ f\Big(-\frac{l_1}{l_3}, y, \frac{\pi}{2}\Big), & l_2 = 0 \end{cases} \tag{4.6}$$

with f the plenacoustic function. Let us consider, as an example, the acoustic scene in figure 4.2. Two OW's lying on the segments $\boldsymbol{p}_A\boldsymbol{p}_B$ and $\boldsymbol{p}_C\boldsymbol{p}_D$ observe a source $\boldsymbol{p}_S$ that is radiating the scene. For simplicity and without loss of generality, we assume the reference frame to be centered on the first camera, with $\boldsymbol{p}_A\boldsymbol{p}_B$ on the y axis. Taking one of the cameras reference frame as the global reference frame is an useful simplification that we will adopt in many occasions from here on. In this case, the chosen camera is referred to as the *reference camera* with the other camera is referred to a the *secondary camera*. As a consequence to this choice, the camera visibility region is strip-shaped, exaclty like in the case presented in 3.2.1. On the other hand, the second OW has a single-wedge shape. Always under the hypotheses of validity of the Radiance Invariance Law, we write the contribution to the plenacoustic image $p_{\boldsymbol{p}_S}(\mathbf{l})$ of the source $\boldsymbol{p}_S$ in the projective ray space as

$$p_{\boldsymbol{p}_S}(\mathbf{l}) = \begin{cases} b_{\boldsymbol{p}_S}\big(\mathrm{sgn}\big(\sin(\theta)\big)\arccos(l_2)\big) & \mathbf{l} \in \mathcal{R}_{\boldsymbol{p}_S} \\ 0 & \text{elsewhere} \end{cases} \tag{4.7}$$

Now that we have seen the features of a joint plenacoustic image, we want to show how it is actually captured and managed in a multiple-camera system.

4.1.3 Acquisition and Fusion of Multiple Plenacoustic Images

Consider the signal acquisition and plenacoustic image computation procedure introduced in 3.2.2. The camera works in its own reference frame, unaware of its position in the world system. We choose to use the same approach also when dealing with multiple cameras. This is beneficial since it allows us to go through the acquisition stage without a prior calibration of the system. As a matter of fact, we will see in 4.3 how we can calibrate the system starting from the individual plenacoustic images that we have captured.

We then proceed as in 3.2.2, considering that each camera lies on the $y^{(i)}$ axis between $y^{(i)} = q_0^{(i)}$ and $y = -q_0^{(i)}$ in its reference frame, acquiring the signal and computing the power plenacoustic image $P^{(i)}(m^{(i)}, q^{(i)}) = |p(m^{(i)}, q^{(i)})|^2$. Again, computing the pseudospectra on a sub-array basis, we can write

$$P^{(i)}(m^{(i)}, q_w^{(i)}) = H_w^{(i)}(\arctan(m)), \tag{4.8}$$

$q_w^{(i)}$ being the y coordinate of the center of the wth sub-array in the local frame of the ith camera. Once computed, we need to map $P^{(i)}(m^{(i)}, q^{(i)})$ into $P^{(i)}(\mathbf{l}^{(i)})$, i.e. transform the local reduced ray spaces to the local projective spaces. This operation corresponds to going from a non-homogeneous to a homogeneous representation. Using the equivalency $m = -l_1/l_2$, $q = -l_3/l_2$ (cfr. section 3.1.1), we write:

$$P^{(i)}(\mathbf{l}^{(i)}) = P^{(i)}\left(-\frac{l_1^{(i)}}{l_2^{(i)}}, \frac{-l_3^{(i)}}{l_2^{(i)}}\right). \tag{4.9}$$

Then, we map the local ray space $P^{(i)}(\mathbf{l}^{(i)})$ onto the global one. Using (4.1):

$$P_i(\mathbf{l}) = P(\mathbf{H}^{(i)^{-T}}\mathbf{l}^{(i)}). \tag{4.10}$$

The global plenacoustic image is then

$$P(\mathbf{l}) = \bigcup_{i=1}^{N} P_i(\mathbf{l}). \tag{4.11}$$

This final equation considers that the visibility regions of the camera are disjointed, i.e. no ray is seen by both cameras simultaneously. In our case this assumption is valid, since the cameras composing our system are purposely non-overlapping (otherwise they would capture the same portion of the source radiation pattern). A real joint plenacoustic image is shown in Figure 4.3.

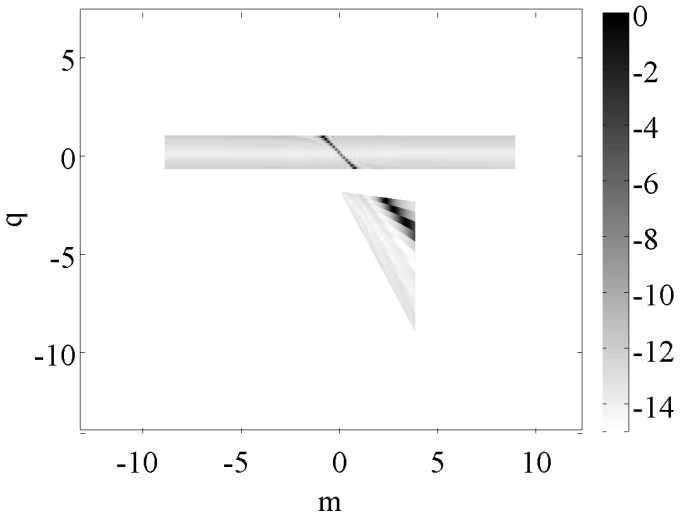


Figure 4.3. Real Joint Plenacoustic Image

The setup is the same of Figure 4.2, where one of the cameras is taken as reference. We call *reference plenacoustic image* the plenacoustic image computed by the reference camera, and *secondary plenacoustic image* the other one. We can see that the distance between peaks in the secondary image is dilated: this is an intrinsic issue of the reduced ray space representation. In the next chapter we explain how we manage the new features of the plenacoustic image in the relevant context of source localization.

4.2 Multiple Camera-Based Source Localization

Source localization is an essential tool to our purpose. In this section we explore how the ability of observing the source from different points of view affects the localization process. Firstly, we explain in detail how the information obtained with the muliple-camera system impacts on the localization method introduced in 3.3, then we illustrate the modifications that we have applied to the method in order to take into account the features of the joint plenacoustic image. For the moment, all the considerations made are still in two dimensions. We extend to the 3D case at the end of this Chapter.

4.2.1 Source Position Estimation Improvement Through Multiple DOA Information

The plenacoustic methods for source localization presented in Chapter 3 rely on the detection and parameters estimation of a line in the plenacoustic image. The (almost) collinear peaks detected on the image, however, result very close to each other and to the origin. This fact affects the estimation of the line's angular coefficient which reflects on an inaccurate estimation of the source distance. What is affected in a lesser way is the DOA estimation. Let us see why. Consider a source in position $\boldsymbol{p}_S = (x, y)^T$ and its dual line in the reduced ray space is

$$\mathcal{I}_{\boldsymbol{p}_S} : q = -xm + y. \tag{4.12}$$

Notice from (4.12) that the angular coefficient of $\mathcal{I}_{\boldsymbol{p}_s}$ is x. We assume that it is estimated with an error ϵ_x: the estimated line, which still intersects the m axis in the point $m_0 = (y/x, 0)^T$, is represented by the equation

$$q = -(x + \epsilon_x)\left(m - \frac{y}{x}\right) \tag{4.13}$$

and has $q-$intercept

$$q_0 = \left(0, (x + \epsilon_x)\frac{y}{x}\right)^T. \tag{4.14}$$

The estimated coordinates of the source are therefore

$$\hat{x} = (x + \epsilon_x) \quad , \quad \hat{y} = (x + \epsilon_x)\frac{y}{x}. \tag{4.15}$$

We can observe that the distance $d = \sqrt{\hat{x}^2 + \hat{y}^2}$ is estimated with an error factor $(x + \epsilon_x)/x$. In fact, if we carry out a few calculations, we obtain

$$d = \sqrt{(x + \epsilon_x)^2 + \left((x + \epsilon_x)\tfrac{y}{x}\right)^2} = \cdots = \frac{x + \epsilon_x}{x}\sqrt{x^2 + y^2}. \tag{4.16}$$

However if no other error is brought by the estimation, the DOA of the source results:

$$\theta = \arctan\left(\frac{\hat{y}}{\hat{x}}\right) = \arctan\left(\frac{(x + \epsilon_x)\frac{y}{x}}{x + \epsilon_x}\right) = \arctan\left(\frac{y}{x}\right) \tag{4.17}$$

which is evidently the correct DOA. It is proven in [11] that the error ϵ_x is almost negligible when the source is close enough to the array (i.e. both DOA and distance are correctly estimated) while it get more and more relevant as the source distance increases.

This is where the secondary camera steps in: a double, independently retrieved information on the DOA can clearly lead to a better estimate of the distance. In plenacoustic terms, the problem reduces to devising a method that exploits the information coming from the two cameras to more accurately estimate the line's parameters in the ray space. Note that the same concept applies to the full ray space: if the homogeneous-coordinates peaks are close to each other, the parameters estimation of the plane they lie upon might not be very accurate. The rays observed by the two cameras form two completely separated clusters. As shown in Figure 4.4, the more this clusters a far apart, the less a peak localization error affects the parameters estimation.

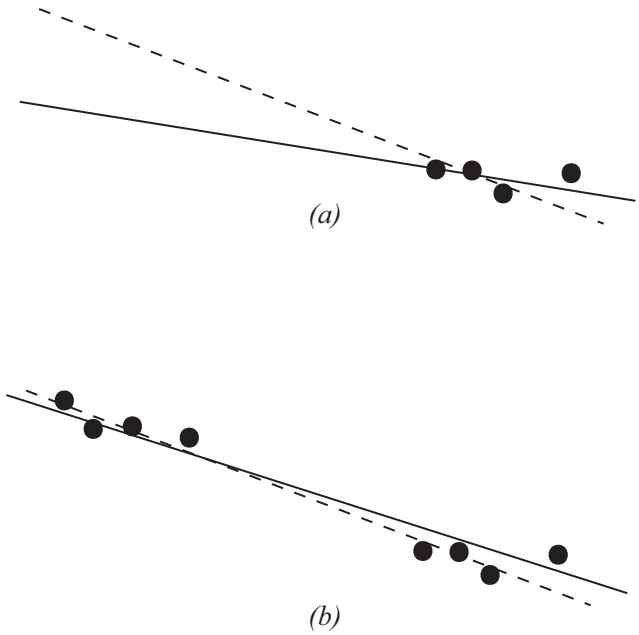
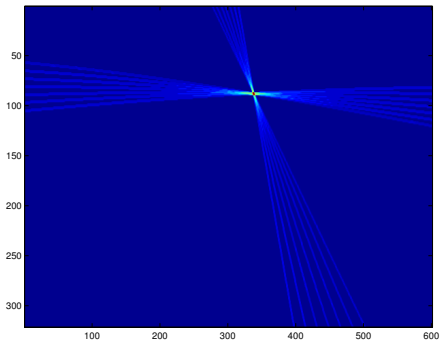


Figure 4.4. Regression error. The dashed line is the one to be estimate through linear regressione, the solid one is the resulting estimation in the case of a) a single cluster, b) two distant clusters

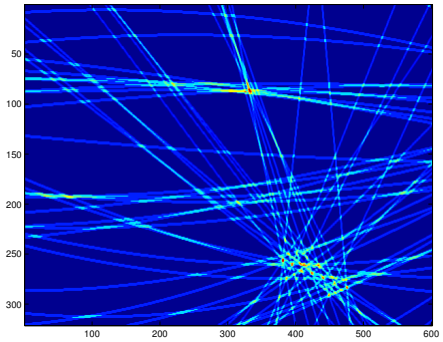
4.2.2 Weighed Hough Transform

We know from section 3.3 that the parameters estimation process of the reduced representation of $\mathcal{I}_{\boldsymbol{p}}$ starts with the Hough Transform. Through some preliminary experiments, we have noticed that the this step might be troublesome in many situations. Let us consider again Figure 4.3. We explained that, when passing to the reduced ray space, the peaks distance is dilated. The peaks farthest from the center of the image generate curves with a steep slope in the Hough-Transform image. The result is show in Figure 4.5(a): thanks to the favorable SNR (10dB), the plenacoustic image shows no peaks due to noise contribution, therefore the curves

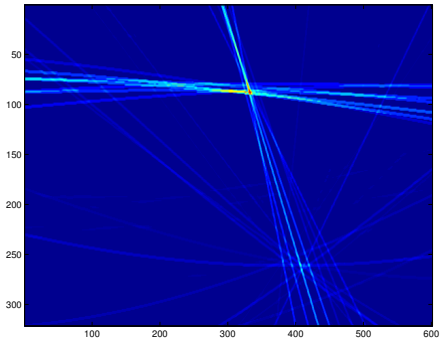
in the parameters-space image are relative only to actual peaks. We can see that the dual curves of the peaks far from the origin exhibit a steep slope and form a cross-like intersection with the curves relative to the other peaks. This is nothing but another way of looking at the concepts introduced in the previous section: the two points of view allow us to better address the estimation process.



(a) Hough Transform, 10dB SNR



(b) Hough Transform, -20dB SNR, unweighed



(c) Hough Transform, -20dB SNR, weighed

Figure 4.5. Hough Transform Images

Now, consider the same scenario but with a low SNR (-10dB): in this case the

Hough-Transform image (shown in Figure 4.5(b)) exhibits several additional curves generated by the noise peaks, some of which are detected in the reference image and some in the transformed secondary image. The result is a tangle that creates numerous intersections and, as a consequence, renders the detection very difficult. We can see that the global maximum of the image might be found at a radiacally wrong location: this error propagates to the clusting algorithm, that erroneously selects noise peaks for the final regression.

Instead of changing the thrersholds for peak detection over the plenacoustic image, we opted for a method that takes into account the magnitude of a peak along with its position. The method defines a set of weights for the curves in the parameters space correponding to the peaks, that takes into account the number of peaks per pseudospectrum (i.e. the likely noisiness) and the ratio between peaks magnitude. The weight $w(v)$ for a peak of magnitude v is computed as

$$w(v) = \frac{w_{max} - w_{min}}{2^o}\left[-\cos\left(\pi\frac{v - v_{min}}{v_{max} - v_{min}} + 1\right)\right]^o + w_{min} \tag{4.18}$$

where o is an order computed as

$$o = \text{round}(0.4N_p). \tag{4.19}$$

N_p is the number of peaks in each pseudospectrum; v_{max}, v_{min} the maximum and minumum magnitude values of the peaks in the image; w_{max}, w_{min} the maximum and minimum weights, set to

$$\begin{aligned} w_{max} &= 10^{(N_p+1)}, \\ w_{min} &= 1. \end{aligned} \tag{4.20}$$

The result of the weighing is shown in Figure 4.5(c).

4.2.3 Localization in the Projective Ray Space

After having detected $\mathcal{I}_{\boldsymbol{p}}$ through the Hough Transform, the next step to the estimation process sees a clustering algorithm and a linear regression on the selected points. However, in this case, we do not consider the inhomogeneous coordinates (m, q) as done in [12]. We choose to consider the homogenous coordinated instead, for a twofold reason. First, as explained in 4.1.2, some rays (specifically those parallel to the y axis, with $l_2 = 0$) tend to infinity as a consequence of the passage to the reduced ray space. This means that they must be discarded when computing the plenacoustic image, resulting in a reduced number of points to feed the regresson with. The second reason is more practical. Consider again the (m, q) reduced ray space. The inhomogeneous coordinates of the points in that space are obtained as $(m, q) = (-l_1/l_2, -l_3/l_2)$. As l_2 gets small, the more a rounding error becomes relevant and is amplified over the resulting reduced coordinates.

Thefore, after detecting the line in the reduced space by means of the Hough Transform, the clustering and the linear regression are accomplished by taking into account the homogeneous coordinates $\mathbf{l} = (l_1, l_2, l_3)^T$ of the rays. The set of rays passing trough a point $\mathbf{p} = (x, y, 1)^T$ is described by the equation $\mathbf{l}^T\mathbf{p} = 0$ or $l_1x + l_2y + l_3 = 0$. We write

$$l_1x + l_2y = -l_3 \tag{4.21}$$

which can be rewritten as

$$\mathbf{l}_{1,2}^T \boldsymbol{p} = -l_3, \quad \mathbf{l}_{1,2} = [l_1, l_2]^T. \tag{4.22}$$

Following the same idea presented in 3.3, equation (3.14) can then be rewritten as:

$$\hat{\boldsymbol{p}} = (\hat{x}, \hat{y})^T = (\mathbf{L}_{1,2}^T \mathbf{L}_{1,2})^{-1} \mathbf{L}_{1,2}^T \mathbf{l}_3, \tag{4.23}$$

where

$$\mathbf{L}_{1,2} = \begin{pmatrix} l_{1_{i_1}} & l_{2_{i_1}} \\ \vdots & \vdots \\ l_{1_{i_N}} & l_{2_{i_N}} \end{pmatrix}, \quad \mathbf{l}_3 = - \begin{pmatrix} l_{3_{i_1}} \\ \vdots \\ l_{3_{i_N}} \end{pmatrix}. \tag{4.24}$$

The estimate $\hat{\boldsymbol{p}}$ represents parameters of the plane $\mathcal{I}_{\boldsymbol{p}_S}$, i.e. the coordinates of the source position. We notice that to perform source localization, we built the joint plenacoustic iamge assuming that the positions of the cameras in the reference frame are known. However, retrieving a precise information of this kind is usually not a trivial task. This is why we devised a plenacoustic method for calibrating the system.

4.3 Autocalibration

In this section we present the autocalibration method that we have developed in the context of this thesis. We have just stated that the position of the cameras composing the system is likely to be unknown a priori. Manual measurement of the rotation matrices $\mathbf{R}^{(i)}$ and of the translation vectors $\mathbf{t}^{(i)}$ would be quite inaccurate. Various methods for calibrating distributed microphones array exist (e.g [16]), but, again, we opted for a solution that exploits the tools we already employ for our goal.

Referring to equation (2.38) presented in section 2.4.2, we can state that a point $\mathbf{p}^{(i)}$ in the ith camera local frame is related to the point $\mathbf{p}_i$ in the global reference frame by:

$$\mathbf{p}_i = \mathbf{H}^{(i)} \mathbf{p}^{(i)}. \tag{4.25}$$

This time, however, rather than providing a general definition of $\mathbf{H}^{(i)}$ we actually define its elements:

$$\mathbf{H}^{(i)} = \begin{pmatrix} \cos(\rho_i) & -\sin(\rho_i) & \Delta x_i \\ \sin(\rho_i) & \cos(\rho_i) & \Delta y_i \\ 0 & 0 & 1 \end{pmatrix} \tag{4.26}$$

where ρ_i and $(\Delta x_i, \Delta y_i)$ are respectively the rotation angle and the x, y translation that lead from the local frame to the global reference frame. The equation (4.25) is still valid if we consider sets of points $\mathbf{P}^{(i)}$ instead of a single one

$$\mathbf{P}_i = \mathbf{H}^{(i)} \mathbf{P}^{(i)} \quad , \tag{4.27}$$

with $\mathbf{P}^{(i)} = [\mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}, \ldots, \mathbf{p}_K^{(i)}]$ and $\mathbf{P}_i = [\mathbf{p}_{1_i}, \mathbf{p}_{2_i}, \ldots, \mathbf{p}_{3_i}]$.

Now, suppose that the cameras perform localization on K different sources independently. Each camera produces a set of estimated positions $\hat{\boldsymbol{p}}_1^{(i)}, \hat{\boldsymbol{p}}_2^{(i)}, \ldots, \hat{\boldsymbol{p}}_K^{(i)}$

relative to its reference system. If we consider the homogeneous coordinates $\hat{\mathbf{p}}^{(i)}$ of the estimations and collect them in a set $\hat{\mathbf{P}}^{(i)}$, we can rewrite (4.27) as

$$\hat{\mathbf{P}}_i = \mathbf{H}^{(i)}\hat{\mathbf{P}}^{(i)}, \tag{4.28}$$

where $\hat{\mathbf{P}}_i$ is the set of (for now, unknown) homogeneous estimates in the reference frame corresponding to the local estimates $\hat{\mathbf{P}}^{(i)}$. The idea is that if we take one of the cameras as reference and decide that its local frame is the reference for the whole system, then the K source position estimates that it has produced are in fact $\hat{\mathbf{P}}_i = [\hat{\boldsymbol{p}}_{1_i}, \hat{\boldsymbol{p}}_{2_i}, \dots, \hat{\boldsymbol{p}}_{K_i}]$. Finding $\mathbf{H}^{(i)}$ from (4.28) is then a LS problem, solved by inverting $\hat{\mathbf{P}}_i$.

However this approach would be quite redundant since it considers $\mathbf{H}^{(i)}$ as a matrix of uncorrelated coefficients. If we look closely at the system

$$\begin{pmatrix} \hat{x}_{1_i} & \hat{x}_{2_i} & \dots & \hat{x}_{K_i} \\ \hat{y}_{1_i} & \hat{y}_{2_i} & \dots & \hat{y}_{K_i} \\ 1 & 1 & \dots & 1 \end{pmatrix} = \begin{pmatrix} \cos(\rho_i) & -\sin(\rho_i) & \Delta x_i \\ \sin(\rho_i) & \cos(\rho_i) & \Delta y_i \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{x}_1^{(i)} & \hat{x}_2^{(i)} & \dots & \hat{x}_K^{(i)} \\ \hat{y}_1^{(i)} & \hat{y}_2^{(i)} & \dots & \hat{y}_K^{(i)} \\ 1 & 1 & \dots & 1 \end{pmatrix} \tag{4.29}$$

it is clear that only 4 parameters need estimating ($\cos\theta$, $\sin\theta$, Δx and Δy). Therefore we write explicitly the system as

$$\begin{cases} \hat{x}_{k_i} = \cos(\rho_i)\hat{x}_k^{(i)} - \sin(\rho)\hat{y}_k^{(i)} + \Delta x_i \\ \hat{y}_{k_i} = \cos(\rho_i)\hat{y}_k^{(i)} + \sin(\rho)\hat{x}_k^{(i)} + \Delta y_i \end{cases}. \tag{4.30}$$

In matrix form, we can write

$$\begin{pmatrix} \hat{x}_{k_i} \\ \hat{y}_{k_i} \end{pmatrix} = \begin{pmatrix} \hat{x}_k^{(i)} & -\hat{y}_k^{(i)} & 1 & 0 \\ \hat{y}_k^{(i)} & \hat{x}_k^{(i)} & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\rho_i) \\ \sin(\rho_i) \\ \Delta x_i \\ \Delta y_i \end{pmatrix} \tag{4.31}$$

or, more compactly:

$$\tilde{\mathbf{P}}' = \tilde{\mathbf{P}}\mathbf{C} \tag{4.32}$$

with $\tilde{\mathbf{P}}' \in \mathbb{R}^{2K}$, $\tilde{\mathbf{P}} \in \mathbb{R}^{2K\times 4}$ and $\mathbf{C} \in \mathbb{R}^4$. Due to dimension requirements, $\tilde{\mathbf{P}}$ must have at least 4 rows, i.e. at least 2 different estimated source positions must be taken into account. Once that requirement is met, $\tilde{\mathbf{P}}$ is invertible (rows and columns are all independent), therefore we are able to calculate all the coefficients we need simply as

$$\mathbf{C} = \tilde{\mathbf{P}}^{-1}\tilde{\mathbf{P}}'. \tag{4.33}$$

Note that, since the position estimations are likely to present errors, even if the system is fully determined with two points it is advisable to use as more points as possibile, so that the linear regression produces a more accurate estimation of the coefficients. In that case the system becomes overdetermined and the regression can be performed as

$$\mathbf{C} = \tilde{\mathbf{P}}^{\dagger}\tilde{\mathbf{P}}' = (\tilde{\mathbf{P}}^T\tilde{\mathbf{P}})^{-1}\tilde{\mathbf{P}}^T\tilde{\mathbf{P}}'. \tag{4.34}$$

Now that we have computed the angle ρ_i and the translation $\Delta x_i, \Delta y_i$ we can build the rotation matrix $\mathbf{R}^{(i)}$ and the translation vector $\mathbf{t}^{(i)}$ that compose the matrix $\mathbf{H}^{(i)}$. In other words, we have all the information we need about the position and pose of the ith camera with respect to the reference camera frame.

4.4 Three-Dimensional Localization From Multiple 2D Localizations

So far, we have seen how we have extendend the single-camera plenacoustic imaging techniques of [12] to a multiple-camera scenario. We have defined the enlarged set of primitives, the relationship between the cameras and the world system, we have explained how to perform a more accurate source localization and how to calibrate the system using the newly introduced plenacoustic concepts. However, this extension is still not enough to fit into the wider context of radiation pattern estimation. This is to the fact that, so far, the plenacoustic camera has been considered to be an uniform linear array working in two dimensions.

The main issue is in the source localization technique: when source localization is performed, the 3D space is cut with the plane on which both the source and the array lie. This means that, assuming that the microphones pickup pattern is omnidirectional, the source needs not be right in front of the array to be detected and localized, but in any case the estimated position is basically always assumed in front of the array.

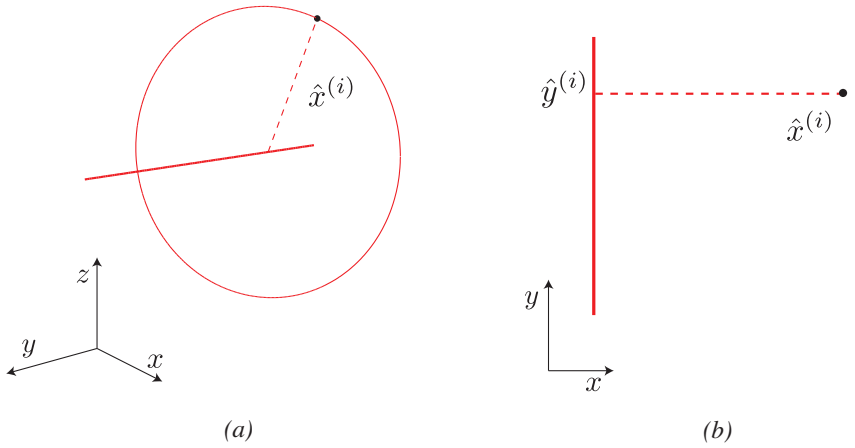


Figure 4.6. Position of a source in 3D space with respect to a ULA (a) and relative localization in 2D (b)

This is an intrinsic issue of the method, due to its reliance on DOA estimation. In the light of this observation, the multiple-camera localization method presented in section 4.2 results to be inapplicable to sources that do not lie on the same plane on which both reference and seconday arrays lie. This means that a source in an arbitrary position in 3D space cannot be correctly localized. Furthermore, as it was proposed in [13], the plenacoustic camera should be a grid of microphones rather than a ULA. We therefore need to make a step further and extend the techniques presented so far to a three-dimensional scenario.

The key aspect of the approach that we adopt is to consider an N by W microphones camera as a set of N vertically aligned ULAs of W microphones each. Our system is composed by two of these cameras (even though there might teoretically be any number), identical in terms of number of microphone and distance between them. Every camera is assumed to be perfectly vertical and to have the same elevation from the ground (so that the corresponding arrays in every camera lie on the same horizontal plane). Starting from this point, we studied a method to derive the 3D source position starting from the 2D estimates provided by arrays composing the cameras.

4.5 3D source position estimation

It is clear that under the perspective suggested in the previous chapters the 2D estimated coordinates $\hat{\boldsymbol{p}}^{(i)} = (\hat{x}^{(i)}, \hat{y}^{(i)})^T$ generated by the ith array are to be treated carfully when passing to 3D space: while the displacement on the y axis is not affected and $\hat{y}^{(i)}$ is valid, $\hat{x}^{(i)}$ actually represents the distance of the source and not its x coordinate. In other words, we are only able to tell that the source lies on a circle of radius $\hat{x}^{(i)}$ centered at $(0, \hat{y}^{(i)})^T$ on the array. Let us consider

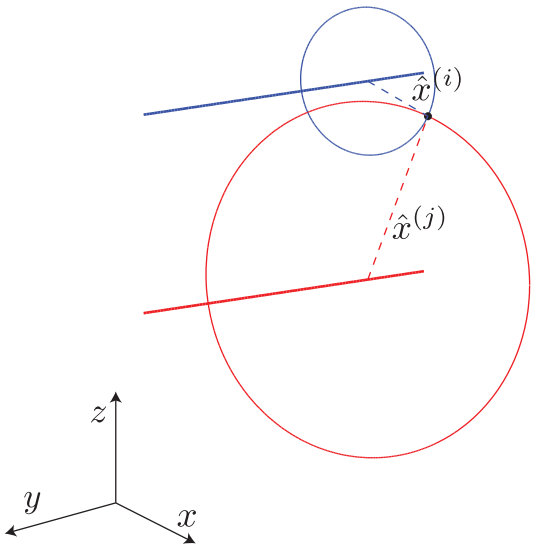


Figure 4.7. Circles of radius $\hat{x}^{(i)}$ centered on the arrays. Example with N=2 arrays.

an $N \times W$ microphones camera as suggested in the previous chapter. Following the aforementioned considerations, we build a geometrical representation of the scene: since we only have one coordinate ($\hat{y}^{(i)}$) and a distance information ($\hat{x}^{(i)}$) for each array composing the camera, the 3D localization problem is transformed in the geometrical problem of finding the intersection of the circles build over that information. We define C as the total number of cameras in the system. The (inhomogeneous) 3D coordinates of the centers of the arrays in the global reference frame are $\boldsymbol{c}_{i,n}$, while $\Theta_{i,n}$ represents their orientation.

4.5.1 Geometrical Representation

First of all we compute the points $\hat{\boldsymbol{p}}_{i,n}$ in the global reference frame corresponding to the points $\hat{\boldsymbol{p}}^{(i,n)}$ in the ith camera local frame. As already explained in the previous chapter, this is done through the isometry:

$$\hat{\boldsymbol{p}}_{i,n} = \mathbf{H}^{(i)}\hat{\boldsymbol{p}}^{(i,n)}. \tag{4.35}$$

We go to 3D space by assigning the z coordinate of the relative array to the estimations, leading to

$$\hat{\boldsymbol{p}}'_{i,n} = \begin{pmatrix} \hat{x}_{i,n} \\ \hat{y}_{i,n} \\ c^z_{i,n} \end{pmatrix}. \tag{4.36}$$

Then we need to build the circles passing through $\hat{\boldsymbol{p}}'_{i,n}$ and centered on the array. In the 3D space a circle is obtained as the intersection of a sphere and a plane. We build the spheres centered on the arrays centers $\boldsymbol{c}_{i,n}$ and have radii

$$d_{i,n} = \|\boldsymbol{c}_{i,n} - \hat{\boldsymbol{p}}'_{i,n}\|. \tag{4.37}$$

As for the planes, they must be normal to the vertical plane on which the arrays lie. If $\Theta_{i,n}$ is the array orientation, the plane coefficients are

$$\begin{aligned} n^x_{i,n} &= \cos(\Theta_{i,n} + \pi/2), \\ n^y_{i,n} &= \sin(\Theta_{i,n} + \pi/2), \\ n^z_{i,n} &= 0. \end{aligned} \tag{4.38}$$

To obtain the circle, for every array i we define the function f^S describing the sphere centered in $\boldsymbol{c}_{i,n}$ and the function f^P describing the plane of coefficients $(n^x_{i,n}, n^y_{i,n}, n^z_{i,n})$ and passing through $\hat{\boldsymbol{p}}'_{i,n}$, then we build the system

$$\begin{cases} f^S: & (x - c^x_{i,n})^2 + (y - c^y_{i,n})^2 + (z - c^z_{i,n})^2 = d^2 \\ f^P: & n^x_{i,n}(x - \hat{x}_{i,n}) + n^y_{i,n}(y - \hat{y}_{i,n}) = 0 \end{cases}. \tag{4.39}$$

The previous equation system represents a single circle built for the nth array of the ith camera. Considering a global system for evey array of evey camera, we are setting up a non-linear problem of finding a common solution to all the equation, i.e. of finding the point where all the circles intesect. The problem can be addressed with the Gauss-Newton algorithm for non-linear least squares approximation.

4.5.2 Gauss-Newton Algorithm

Our goal is to find a solution $\bar{\boldsymbol{p}}$ that minimizes the error $\|\mathbf{r}(\bar{\boldsymbol{p}})\|^2 = \|\mathbf{f} - \mathbf{d}\|^2$, where

$$\mathbf{f} = \begin{pmatrix} f^S(\boldsymbol{c}_{1,1}) \\ f^P(\hat{\boldsymbol{p}}_{1,1}) \\ \vdots \\ f^S(\boldsymbol{c}_{C,N}) \\ f^P(\hat{\boldsymbol{p}}_{C,N}) \end{pmatrix} \quad \mathbf{d} = \begin{pmatrix} d^2_{1,1} \\ 0 \\ \vdots \\ d^2_{C,N} \\ 0 \end{pmatrix} \tag{4.40}$$

The solution $\bar{\boldsymbol{p}}$ is the 3-vector representing the inhomogeneous cartesian coordinates of the source.

We choose the estimate $\hat{\boldsymbol{p}}'_{i,n}$ with minimum norm as the initial solution $\bar{\boldsymbol{p}}^{(0)}$ to feed the algorithm with

$$\bar{\boldsymbol{p}}^{(0)} = \arg\min_{\bar{\boldsymbol{p}}'_{i,n}}(\|\bar{\boldsymbol{p}}'_{i,n}\|). \tag{4.41}$$

This is due to the fact that this particular estimate is more likely to have x and y coordinates close to the actual 3D position. Inizitalizing $\bar{\boldsymbol{p}}^{(k)} = \bar{\boldsymbol{p}}^{(0)}$, the Gauss-Newton algorithm can now be implemented as follows:

while $\|\bar{\boldsymbol{p}}^{(k+1)} - \bar{\boldsymbol{p}}^{(k)}\| > t_c$ **do**

- Linearize $\mathbf{r}$ near current iterate $\bar{\boldsymbol{p}}^{(k)}$:

$$\mathbf{r}(\bar{\boldsymbol{p}}) \approx \mathbf{r}(\bar{\boldsymbol{p}}^{(k)}) + \mathbf{D}(\bar{\boldsymbol{p}}^{(k)})(\bar{\boldsymbol{p}} - \bar{\boldsymbol{p}}^{(k)}) \tag{4.42}$$

  where $\mathbf{D}$ is the Jacobian whose elements are $\mathbf{D}|_{ij} = \delta\mathbf{r}_i/\delta\bar{\boldsymbol{p}}_j$;

- compute next iteration $\bar{\boldsymbol{p}}^{(k+1)}$ as the Least-Squares solution of $||A^{(k)T}\bar{\boldsymbol{p}} - b^{(k)}||^2$:

$$\bar{\boldsymbol{p}}^{(k+1)} \leftarrow (A^{(k)T}A^{(k)})^{-1}A^{(k)T}b^{(k)} \tag{4.43}$$

  where: $A^{(k)} = \mathbf{D}(\bar{\boldsymbol{p}}^{(k)})$ and $b^{(k)} = \mathbf{D}(\bar{\boldsymbol{p}}^{(k)})\bar{\boldsymbol{p}}^{(k)} - [\mathbf{f}(\bar{\boldsymbol{p}}^{(k)}) - \mathbf{d}]$;

- Increment counter

$$k \leftarrow k + 1$$

end

Algorithm 1: Gauss-Newton Algorithm

Note that in our case the Jacobian matrix $\mathbf{D}$ is defined as:

$$\mathbf{D} = \begin{pmatrix} \frac{\delta f^S(\boldsymbol{c}_{1,1})}{\delta\bar{x}} & \frac{\delta f^S(\boldsymbol{c}_{1,1})}{\delta\bar{y}} & \frac{\delta f^S(\boldsymbol{c}_{1,1})}{\delta\bar{z}} \\ \frac{\delta f^P(\hat{\boldsymbol{p}}_{1,1})}{\delta\bar{x}} & \frac{\delta f^P(\hat{\boldsymbol{p}}_{1,1})}{\delta\bar{y}} & \frac{\delta f^P(\hat{\boldsymbol{p}}_{1,1})}{\delta\bar{z}} \\ \vdots & \vdots & \vdots \\ \frac{\delta f^S(\boldsymbol{c}_{C,N})}{\delta\bar{x}} & \frac{\delta f^S(\boldsymbol{c}_{C,N})}{\delta\bar{y}} & \frac{\delta f^S(\boldsymbol{c}_{C,N})}{\delta\bar{z}} \\ \frac{\delta f^P(\hat{\boldsymbol{p}}_{C,N})}{\delta\bar{x}} & \frac{\delta f^P(\hat{\boldsymbol{p}}_{C,N})}{\delta\bar{y}} & \frac{\delta f^P(\hat{\boldsymbol{p}}_{C,N})}{\delta\bar{z}} \end{pmatrix} \tag{4.44}$$

Since the algorithm does not guarentee convergence, the threshold t_c must be set in advance in order to obtain an estimate in a finite number K of iteration. The outupt value $\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}^{(K)}$ is the 3D source position estimate we are looking for.

4.6 Conclusions

In this chapter, we have covered the core aspect of our studies, that is the extension of plenacoustic analysis to a multiple-view model, presenting the key concepts and introducing the new formal notation. Part of this study is the multiple-camera source localization method, that improves its single-camera counterpart by exploting the ability of observing the source from different points of view. Furthermore, the autocalibration method that we have devised allows us to immediately report the study to a real context: in our case, being able to calibrate the system accurately is essential, since both the source localization and radiation pattern estimation methods require that the relative position of cameras is known. The final step towards a completely plenacoustic analysis-based radiation pattern estimation system is the 3D source localization technique presented in the last section of this Chapter. Once all of these aspects have been addressed, we can move forward and show how the new concepts integrate to the system first introduced in [13]: in the next Chapter we see how this integration is done and how the resulting system benefits in terms of compactness.

Chapter 5

Radiation Pattern Estimation

In the previous chapters we introduced the plenacoustic tools already available in the literature as well as the novel approaches that we have studied with the aim of improving the technique for radiation pattern estimation devised in [13]. Multiple-camera plenacoustic imaging enables a more precise analysis of the soundfield, allowing us to perform a more reliable source localization. Thanks to the method presented in Section 4.5 we are able to estimate the 3D position of a source integrating 2D plenacoustic techniques with some geometric considerations.
In this chapter we see how the newly devised methods can be applied to the multiple-camera radiation pattern system that we intend to investigate in this thesis. We explain how the technique of [13] is reformulated by taking into account the fully plenacoustic approach. Since the focus of our work is mainly on the multiple-view extension of plenacoustic imaging, we here refer to a slightly more basic system: the source is considered controlled in position and in terms of signal emission (a loudspeaker or an instrument in fixed positions) and the aspects related to frame selection and dynamic compensation that are included in [13] are not covered, as they are not very relevant to our problem. The idea is that the source emits a signal in a fixed position, then it is rotated arbitrarily and possibily into another position before emitting another signal. In this sense, we explore a first approximation of an actual moving source. We explain how the position over the acquisitions can be estimated with the methods we have devised and presented in the previous Chapter. Also, we explain how this information can be combined with the orientation information provided by an external device (a gyroscope). After setting the basic notation that will be used throughout the chapter, we explain in a progressive fashion how the whole acquired signal is managed and analyzed.

5.1 General Setup

We consider a two-cameras setup where each camera is composed by N arrays of W microphones stacked to form a rectangular grid (where the vertical and horizontal distance between microphones is equal). We identify one of the two cameras to be the reference. The origin of the whole reference system is the center of the bottom array of the reference camera and is defined by the inertial unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. In the global reference system, we consider the inhomogeneous 3D coordinates of the

centers $\boldsymbol{c}_{i,n}$ of the individual arrays as defined in Section 4.5. We here also define the global 3D position of each microphone as:

$$\boldsymbol{m}_{i,n,w} = \begin{pmatrix} m^x_{i,n,w} \\ m^y_{i,n,w} \\ m^z_{i,n,w} \end{pmatrix} \tag{5.1}$$

where $i = 1, 2$ is the camera index; $n = 1, \ldots, N$ the array index; $w = 1, \ldots, W$ the microphone index. Each microphone acquires synchronously the signal $s_{i,n,w}(t)$.

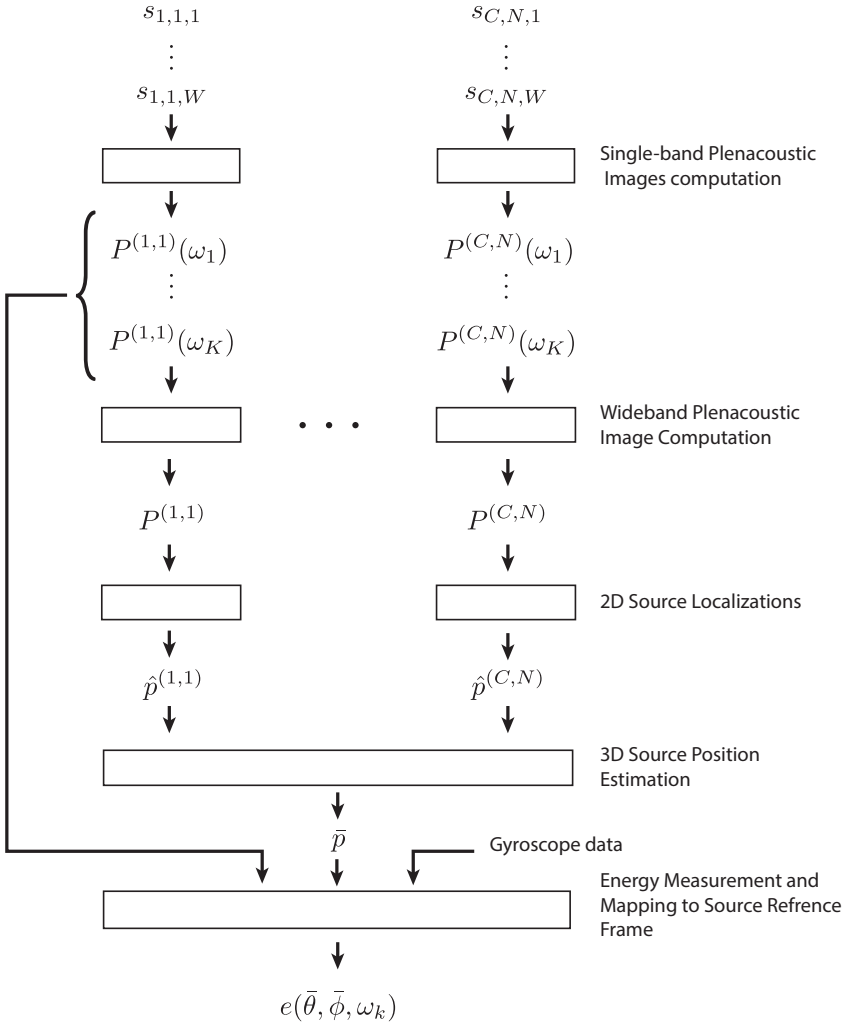


Figure 5.1. General Setup

5.2 Source Position and Orientation Estimation

As a first step, we compute the pseudospectrum with each array n independently through a MVDR beamformer (cfr. Appendix A) on the signals $s_{i,n,w}(t)$. 2D source localization is then performed as in 3.3 to obtain the individual 2D source position estimates:

$$\hat{\mathbf{P}}_i = [\hat{\mathbf{p}}_{i,1}, \ldots, \hat{\mathbf{p}}_{i,N}]. \tag{5.2}$$

Using the method described in Section 4.5 we derive the global 3D source position estimate

$$\bar{\boldsymbol{p}} = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix}. \tag{5.3}$$

The relevant aspect of this steps is that the plenacoustic images here computed are also used during the energy measurement step. While the position is estimated through plenacoustic methods, we still need an external device for determining the source orientation with respect to our system.

The orientation of the source is obtained as a combination of its rotation about three axes. The value of these rotations, respectively called *pitch* (ψ), *yaw* (φ) and *roll* (μ), may be obtained through a *gyroscope*. The details on this kind of device and on how to manage the data it outputs is be provided in Section 6.2. For the time being we only need to take for granted that (ψ, φ, μ) are known for every t and that their value is 0 when the source is aligned with the whole system's reference frame. We therefore compute the global orientation by rotating the source's reference frame by the obtained angles. To do so, initially assuming that the source is located at the origin of our reference system, we apply three successive algebraeic rotations to the inertial vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$. These rotations about the inertial axes x, y, z are respectively defined as:

$$\mathbf{R}_\psi = \begin{pmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{pmatrix}, \tag{5.4}$$

$$\mathbf{R}_\varphi = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 1 & 1 & 0 \end{pmatrix}, \tag{5.5}$$

$$\mathbf{R}_\mu = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\mu & \sin\mu \\ 0 & -\sin\mu & \cos\mu \end{pmatrix}. \tag{5.6}$$

Combined in the presented order, they provide the complete rotation matrix $\mathbf{R} = \mathbf{R}_\psi \mathbf{R}_\varphi \mathbf{R}_\mu$, which can now be applied to the intertial unit vectors leading to

$$\mathbf{i}' = \mathbf{R}\mathbf{i}, \qquad \mathbf{j}' = \mathbf{R}\mathbf{j}, \qquad \mathbf{k}' = \mathbf{R}\mathbf{k}. \tag{5.7}$$

Lastly, we translate by the source position estimate $\bar{\mathbf{x}}$ to obtain the complete rototranslated source reference frame described by the inertial unit vectors:

$$\bar{\mathbf{i}} = \mathbf{i}' + \bar{\mathbf{x}}, \qquad \bar{\mathbf{j}} = \mathbf{j}' + \bar{\mathbf{x}}, \qquad \bar{\mathbf{k}} = \mathbf{k}' + \bar{\mathbf{x}}. \tag{5.8}$$

5.3 Energy Measurement

Once all the information about source position and orientation is obtained, we can proceed with measuring the energy it radiates in all those directions that are visible to our plenacoustic cameras. At first we see how to retrieve this information from the plenacoustic images already acquired for source localization. Then, by mapping the energy information to the source reference frame and compensating for the source distance we can derive the global source radiance pattern.

5.3.1 Energy Measurement from Plenacoustic Images

What we have at this point is a set of plenacoustic images, one for each ULA. We here consider K single-band plenacoustic images rather than the wideband version used for localization (cfr. 3.2), so that for a given $\bar{k}$ we have a relative image $P^{(i,n)}(\theta, \bar{\omega}_k)$. As explained in 3.2.1, the plenacoustic image is composed by the pseudospectra $\mathbf{h}_w(\theta, \omega_k)$ computed for each sub-array w. We know that the peak in the pseudospectrum is the value of the energy coming from the direction $\bar{\theta}$ of the source. Therefore, considering the wth sub-array of the nth array in camera i, we define the ray originating at the source location and directed to the central microphone $\boldsymbol{m}_w$ as:

$$\mathbf{v}_{i,n,w} = \boldsymbol{m}_{i,n,w} - \bar{\mathbf{x}} \quad \text{with} \quad |\mathbf{v}_{i,n,w}| = r_{i,n,w}, \tag{5.9}$$

$r_{i,n,w}$ being the distance between the source and the central microphone. We define the energy associated to this ray as

$$e(\mathbf{v}_{i,n,w}, \omega_k) = P^{(i,n,w)}(\bar{\theta}, \omega_k) \tag{5.10}$$

where $P^{(i,n,w)}(\bar{\theta}, \omega_k)$ is the kth band plenacoustic image computed at the sub-array w of array (i, n). Each sub-array is then able to observe a different ray coming from the source location and provide information about its energy. This means that the two $N \times W$ microphones grids are able to observe $2 \times N \times (W - J + 1)$ rays for each acquisition, with J the number of microphone in a sub-array. We can see how the extension to a multiple-camera setup is beneficial to our problem: thanks to the acoustic-rays based representation of the soundfield peculiar to plenacoustic imaging, we are able to extend our OW to multiple separate points of view, in a straighforward way and with no modification to the underlying model. Furthermore, in the case of a real musical performace, this addition allows the musician to move in a natural way, without the need of eccessively rotating about himself in order to present all the pattern to the plenacoustic cameras.

5.3.2 Mapping to Source Reference Frame

In order to obtain a radiation pattern, the information gathered so far must be expressed in the source reference frame. We have a set of rays $\mathbf{v}_{i,n,w}$, one for each sub-array, associated to an energy measure $e(\mathbf{v}_{i,n,w})$. If we define the wth microphone reference frame with the three unit vectors $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$, we can write (dropping the indices for the sake of simplicity of notation)

$$\mathbf{v} = \hat{v}_x \hat{\mathbf{i}} + \hat{v}_y \hat{\mathbf{j}} + \hat{v}_z \hat{\mathbf{k}}. \tag{5.11}$$

We perform the reference frame change with

$$\bar{\mathbf{v}} = \begin{pmatrix} \bar{v}_{\bar{x}} \\ \bar{v}_{\bar{y}} \\ \bar{v}_{\bar{z}} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{i}} \cdot \bar{\mathbf{i}} & \hat{\mathbf{j}} \cdot \bar{\mathbf{i}} & \hat{\mathbf{k}} \cdot \bar{\mathbf{i}} \\ \hat{\mathbf{i}} \cdot \bar{\mathbf{j}} & \hat{\mathbf{j}} \cdot \bar{\mathbf{j}} & \hat{\mathbf{k}} \cdot \bar{\mathbf{j}} \\ \hat{\mathbf{i}} \cdot \bar{\mathbf{k}} & \hat{\mathbf{j}} \cdot \bar{\mathbf{k}} & \hat{\mathbf{k}} \cdot \bar{\mathbf{k}} \end{pmatrix} \begin{pmatrix} \hat{v}_x \\ \hat{v}_y \\ \hat{v}_z \end{pmatrix} \tag{5.12}$$

where $(\bar{\mathbf{i}}, \bar{\mathbf{j}}, \bar{\mathbf{k}})$ are the source reference frame's unit vectors computed in 5.2. The above equation simplifies in

$$\begin{pmatrix} \bar{v}_{\bar{x}} \\ \bar{v}_{\bar{y}} \\ \bar{v}_{\bar{z}} \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{i}}_{\bar{x}} & \bar{\mathbf{i}}_{\bar{y}} & \bar{\mathbf{i}}_{\bar{z}} \\ \bar{\mathbf{j}}_{\bar{x}} & \bar{\mathbf{j}}_{\bar{y}} & \bar{\mathbf{j}}_{\bar{z}} \\ \bar{\mathbf{k}}_{\bar{x}} & \bar{\mathbf{k}}_{\bar{y}} & \bar{\mathbf{k}}_{\bar{z}} \end{pmatrix} \begin{pmatrix} \hat{v}_x \\ \hat{v}_y \\ \hat{v}_z \end{pmatrix}. \tag{5.13}$$

We then convert the vector into spherical coordinates in order to map the energy information on a sphere centered in $\bar{\mathbf{x}}$:

$$\bar{\theta} = \arctan\left(\frac{\bar{v}_{\bar{y}}}{\bar{v}_{\bar{x}}}\right), \tag{5.14}$$

$$\bar{\phi} = \arctan\left(\frac{\bar{v}_{\bar{z}}}{\sqrt{\bar{v}_{\bar{x}}^2 + \bar{v}_{\bar{y}}^2}}\right), \tag{5.15}$$

$$r = \sqrt{\bar{v}_{\bar{x}}^2 + \bar{v}_{\bar{y}}^2 + \bar{v}_{\bar{z}}^2}. \tag{5.16}$$

Finally, we define the energy associated to these coordinates as:

$$e(\bar{\theta}, \bar{\phi}, r, \omega_k) = e(\mathbf{v}_{i,n,w}, \omega_k). \tag{5.17}$$

The energy values $e(\bar{\theta}, \bar{\phi}, r, \omega_k)$, though, do not yet constitute a radiation pattern for a twofold reason: they are mapped onto difference distances r from the cameras and they do not take into account dynamic variations of the signal. In the next section we refine the definition of the energy value e taking into account these two factors in successive steps.

5.3.3 Distance and Dynamic Compensation

According to the Rayleigh's Integral (2.4), a wave propagating in free field is subjected to an amplitude attenuation of $1/r$. Due to the multiple-camera setup and the planar distribution of the microphones in each camera, sub-arrays are at different distances r from the source. We therefore need to compensate these different losses by multiplying the energy values $e(\bar{\theta}, \bar{\phi}, r, \omega_k)$ by the squared distances r^2. Doing so, we are implicitly mapping all the values extracted with the plenacoustic cameras on a spere of radius 1 m, thus obtaining a directivity pattern. For each sub-array w, at each frequency band k, the energy value is defined by:

$$e(\bar{\theta}, \bar{\phi}, \omega_k) = e(\mathbf{v}_{i,n,w}, \omega_k) r^2. \tag{5.18}$$

Now, we need to take into account dynamic variations of the signal. This step is of particular importance when dealing with musical instrument. The instrument can be excited with different intensities, even if it is in a fixed positions. In the case of musical instruments we therefore resort to a reference signal whose energy can be used to compensate these dynamic variations (in chapter 6 we explain how to obtain this signal).
We calculate the energy of the reference signal $s_r(t)$ as:

$$e_r = \sum_{t=1}^{T} s_r(t)^2, \tag{5.19}$$

then divide the extracted pseudospectrum value with it, to compensate the dynamic changes, leading to

$$e(\bar{\theta}, \bar{\phi}, \omega_k) = e(\mathbf{v}_{i,n,w}, \omega_k)\frac{r^2}{e_r}, \tag{5.20}$$

which is the final directivity pattern value relative to sperical coordinates $(\bar{\theta}, \bar{\phi})$

5.4 Fusion of Multiple Measurements

We have seen how to measure the energy radiated by the source in every direction. If a portion of the radiance is observed more than once over the acquisitions, we have more than one measure of the energy associated to spherical angels $(\bar{\theta}, \bar{\omega})$ and frequency bin ω_k. We can compute an average as

$$e(\bar{\theta}, \bar{\phi}, \omega_k) = \frac{1}{D(\bar{\theta}, \bar{\phi}, \omega_k)} \sum_{d=1}^{D(\bar{\theta}, \bar{\phi}, \omega_k)} e_d(\bar{\theta}, \bar{\phi}, \omega_k), \tag{5.21}$$

where $D(\bar{\theta}, \bar{\phi}, \omega_k)$ is the total number of measures for $e(\bar{\theta}, \bar{\phi}, \omega_k)$ and d the index of the single measure.

5.5 Conclusions

In this chapter we discussed how multiple-view plenacoustic imaging applies to the problem of radiation pattern estimation and better exploits its potential. The system is able to work with a very limited need of external devices: as a matter of fact, except for the gyroscope that is necessary to track the source orientation over time, it exploits the same acquisition data to both localize the source and measure the energy coming from it at a number of different angles. The plenacoustic imaging techniques can provide all the information needed at all the main steps of the process, rendering the analysis more coherent and self-contained.

In the next Chapter we present some simulated and experimental data that show how the various components of the system perform.

Chapter 6

Implementation

After presenting our studies on multiple-camera plenacoustic imaging and the application of these studies to a radiation pattern estimation system, we explain how our system is implemented. At first, we present the result of a series of simulations that we have set up in order to show the effectiveness of the techniques that we have devised. Then we briefly present the hardware and software components of our measurement system. At the end of the chapter we also present a series of experimental results to confirm the validity of our system. These results are referred to the acquisition of a loudspeaker-generated signal, and show how the radiation pattern is estimated.

6.1 Simulations

The results presented in this section is obtained through a series of *Matlab* scripts that simulate sound sources and plenacoustic cameras and process this data through the same algorithms that are used in the context of actual acquisitions. We consider ULAs composed of 9 microphones spaced by 0.06m, for a total array length of 0.48m. First, we show the performance of the autocalibration method, then we cover the source localization methods (2D and 3D).

6.1.1 Autocalibration

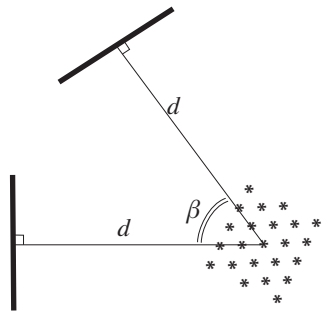
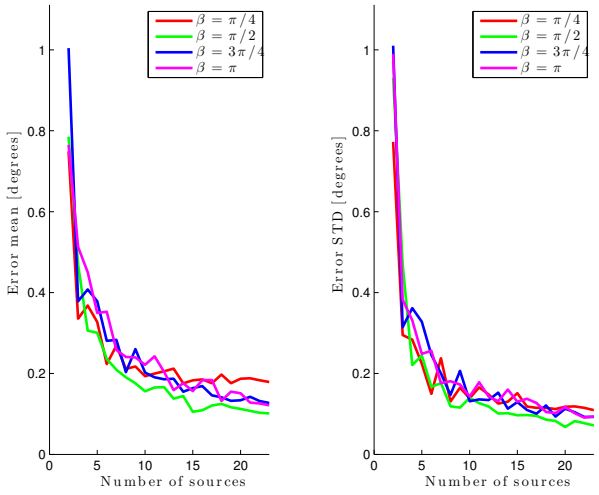
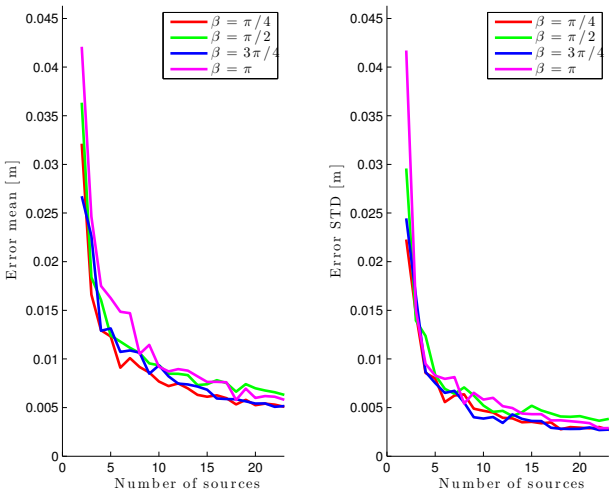


Figure 6.1. Autocalibration simulation setup: the stars represent the sources lattice, while the two bars represent the ULAs

We evaluated the autocalibration error as the mean orientation and center position estimation error. In figure 6.1 we show the setup of the simualtion: 23 sources arranged on a lattice are positioned at distance d from both the reference and the secondary camera. A white noise signal is generated by one source at a time with a 10dB SNR, and the source is localized independently by both cameras. A total of $I = 50$ acquisitions and localizations are performed for every source. The autocalibration method presented in Section 4.3 is applied by taking into account a number $N = 2, \ldots, 23$ of localizations, chosen randomly between the available ones iteratively for I times.



(a) Orientation Estimation Error At d=1.5m

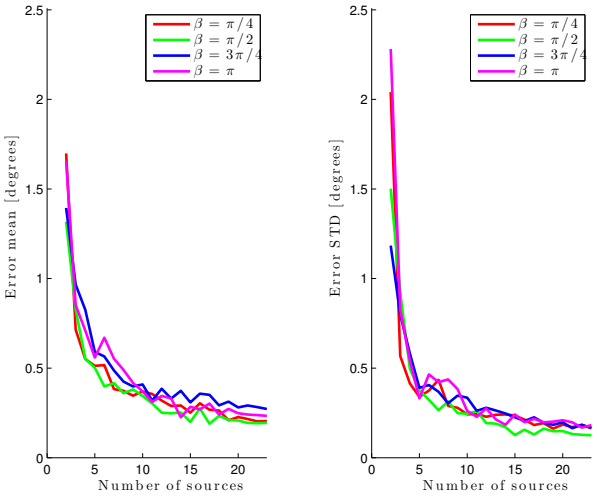


(b) Center Estimation Error At d=1.5m

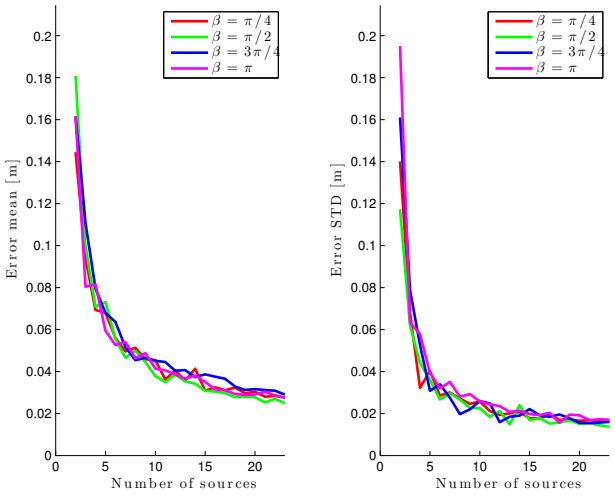
Figure 6.2. Autocalibration errors at d=1.5m

Then, the angle β is varied and the process repeated. Once in a first simulation

all angles of interest are covered, in a second simulation the distance d is varied as well and the whole process is repeated from the beginning. In Figures 6.2, 6.3 and 6.4 we present the results for three values of d ($d = 1\text{m}, 2\text{m}, 3\text{m}$), with β varying from $\pi/4$ to π and N from 2 to 13. We observe that the error is stable enough from about 10 sources for all distance values. In general, the error increases with the distance, due to the fact that the perfomanes of the single-camera based localization method decrease (as explained in 4.2.1). Note that the irregualar shape of the curves is to



(a) Orientation Estimation Error At d=2.5m

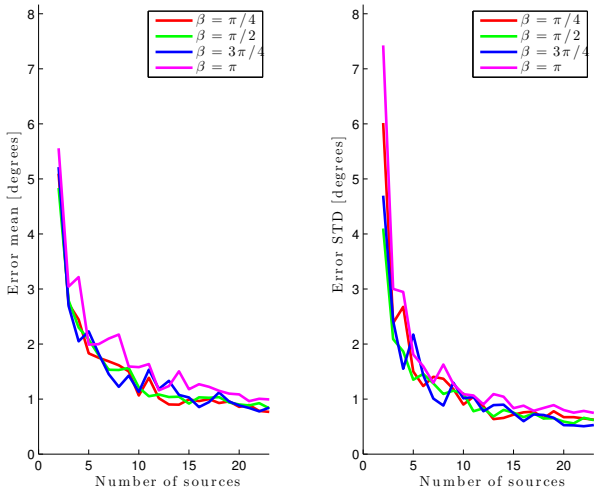


(b) Center Estimation Error At d=2.5m

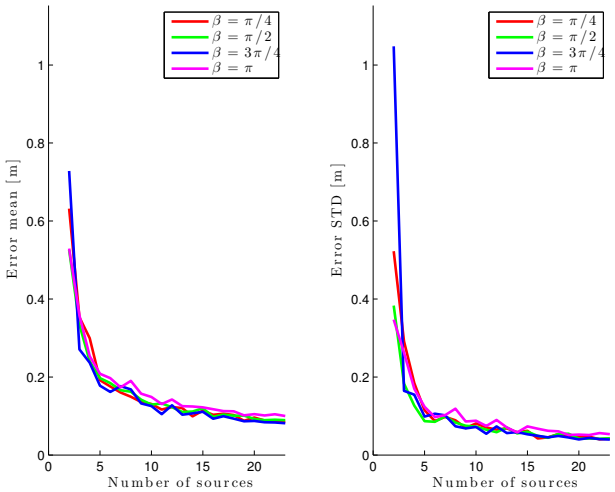
Figure 6.3. Autocalibration errors at d=2.5m

be ascribed to the error computation method explained above: the N number of single-array localizations is chosen randomly between the the 23 ones available. This means that if a noisy localization is included, the mean error (and STD) increases

slightly.



(a) Orientation Estimation Error At d=3.5m



(b) Center Estimation Error At d=3.5m

Figure 6.4. Autocalibration errors at d=3.5m

6.1.2 Multiple-Camera Source Localization

We here consider that the two cameras are perfectly calibrated, i.e. their relative pose and position is known with no error. Again, each camera is composed by 9 microphones and is 0.48m long. First of all, we present the localization error of a simulated source at location $\boldsymbol{x} = [1\text{m}, 0\text{m}]^{\text{T}}$ that generates a with noise signal. We evaluate joint-localization error as a function of the SNR, varying from -10dB to 10dB. The main goal of this simulation is to evaluate the localization error compared

to a single-camera case, where a camera of 16 microphones performs localization on a source at the same location. Note that the 16 microphones camera is 1m long, so that the distance between sensors remains approximately 0.07m. In this sense, the single camera presents a double number of sub-arrays that is double with respect to the two smaller cameras, i.e. a double number of pseudospectra compose its plenacoustic image. The results are shown in Figure 6.5.Firstly, as explained in 4.2.3, in the multiple-camera case is necessary to take into account the homogeneous coordinates of the peaks in the plenacoustic image. In this simulation we clearly see how the regression over inhomogeneous coordinates would be significantly inaccurate. Secondly, we here confirm that the multiple-camera localization error is always less than the single double-length camera, which is in line with what was stated in Chapter 4. We also evaluate the performance in relation to the distance of the source: in figure 6.6 we can see that, always following what was stated in 4.2.1, the multiple-camera techniques are far more robust that the single-camera ones. This is due to the fact that by combining two independent DOA information we are able to obtain a better distance estimation as well.

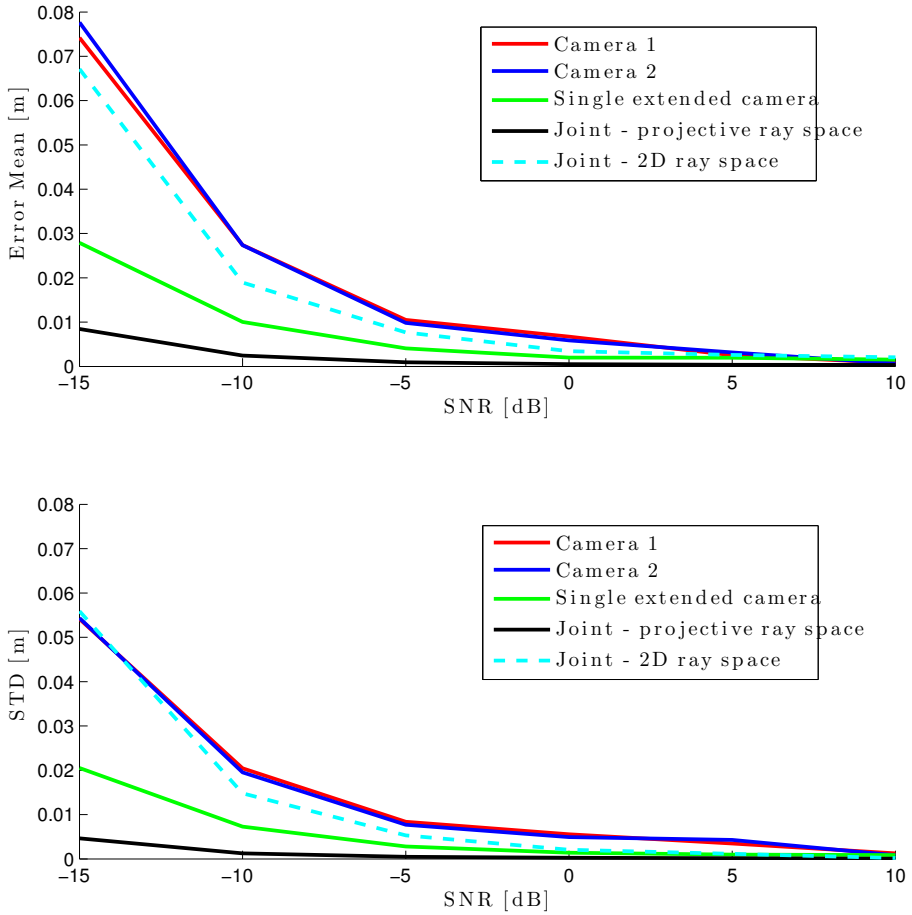


Figure 6.5. Multiple-Camera Localization Error Vs. SNR

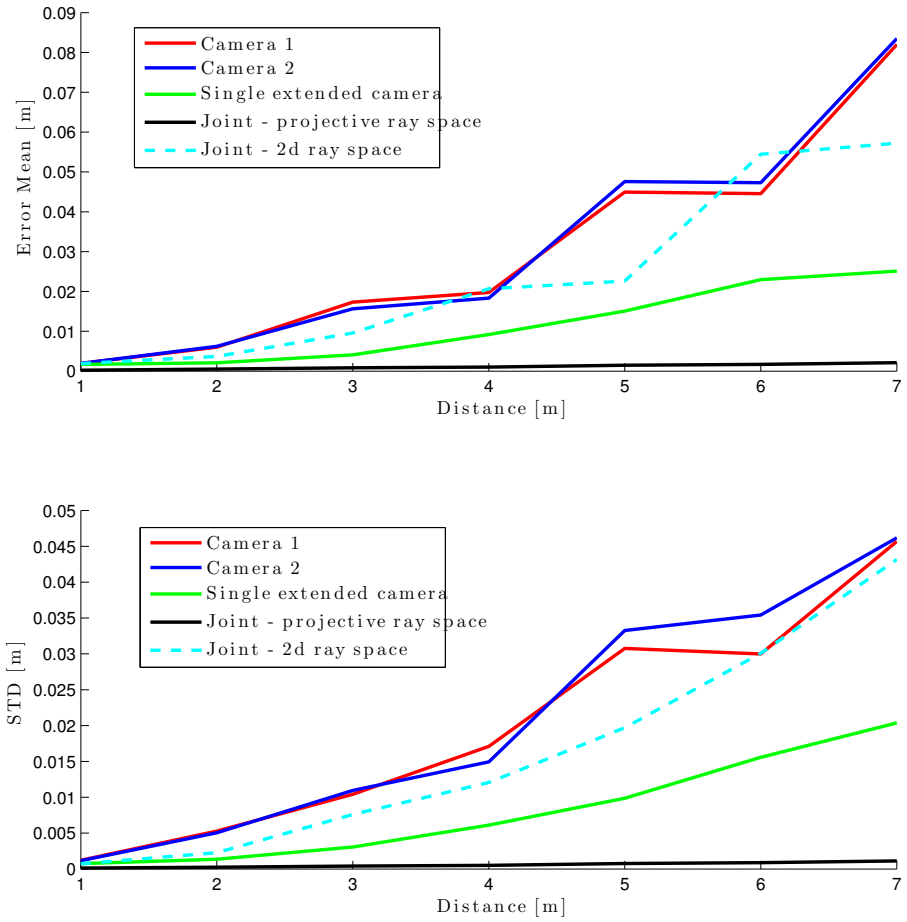


Figure 6.6. Multiple-Camera Localization Error Vs. Distance

6.1.3 3D Source Position Estimation

Also in this case, we consider that the cameras are correctly calibrated. The cameras are two 4×9 microphones grids, with sensors spaced by 0.07m horizontally and vertically. The bottom arrays of both cameras are at 0m. The 3D source position estimation method presented in Section 4.5 relies on the invididual 2D localizations provided by each array composing the microphone-grid cameras. For this reason, we simulate a source at location $\boldsymbol{x} = [0.2, 1.1, 0.11]^T$ that emits a white noise signal, then we simulate acquisition and localization on each array independently and evaluate the error on the 3D source position estimation at varying SNR and values of β (cfr. Figure 6.1) . We can see in Figure 6.7 that the error decreases as the SNR increases and that it is not significantly dependent on β.

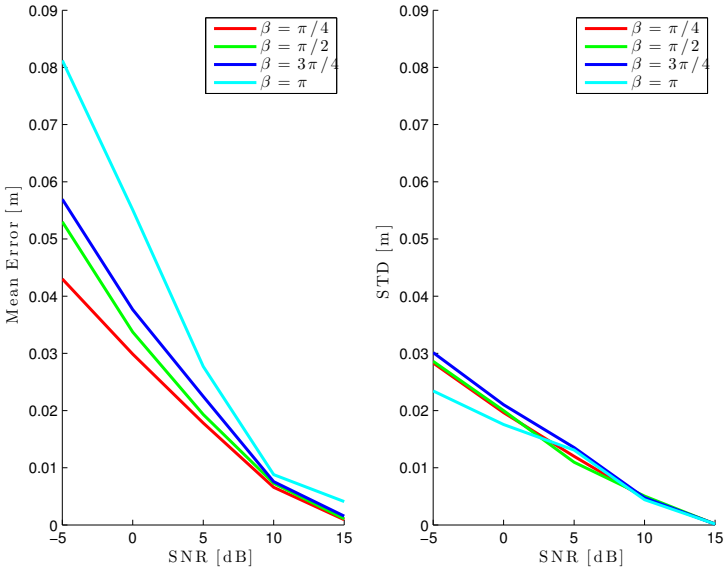


Figure 6.7. 3D localization errors at varying SNR

6.2 Setup

6.2.1 Hardware

The system is composed by two plenacoustic cameras that we built specifically for this application from wooden boards. A total of 64 *Beyerdynamic MM1* condenser measurement microphones is arranged in two 4×8 microphones regular grids, with each element spaced vertically and horizontally by 0.07m. The grids are obtained from two wooden boards drilled to accomodate the microphones, supported by tripod stands that are fixed so that the bottom array of each camera is at 1.46m from the ground.

All microphones are connected to eight *Aphex 188* 8-channel preamplifiers whose output is routed to two *Apogee Symphony I/O* 32-channel A/D audio interfaces. A Mac Pro is used for the acquisition process.

We anticipated in Section 5.2 that, in order to track the varying orientation of the source, we need a gyroscope. The one we use is *Phidgets Spatial 3/3/3 1044*, which provides three-axes angular rate via USB.

The reference microphone is a generic piezoelectric microphone, attached close to the bridge of the violin. Notice that the reference microphone, as explained in Section 5.3.3, is used when dealing with an actual insturment and only provides a reference energy signal used to compensate possible dynamic variations.

Technical specifications of all hardware components are provided in Appendix B

6.2.2 Software

The software side of the system is composed of two applications, specifically implemented for this work. The first one, devoted to the acquisition process, is

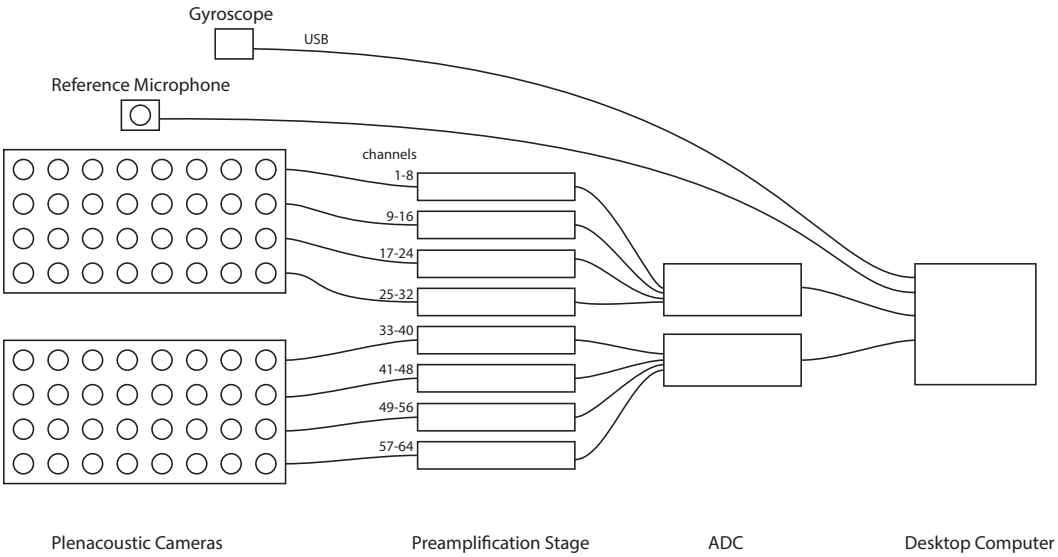


Figure 6.8. Block scheme of the hardware system

developed in *Obj-C* and features a graphical interface (Figure 6.9). This application exploits the *Portaudio* framework to acquire the 64 microphone signals outputted by the *Symphony I/O* interface, along with the reference microphone signal and the gyroscope data. The microphones signals are sampled at 44.1KHz and quantized at 16bits precision. A buffer of 512 samples per each channel is saved to a raw binary file every 0.0116s (44100samples/s $\times$ [512samples/buffer]$^{-1}$). When the buffer is saved, orientation data is retrieved from the gyroscope. This data is provided in the form of 3-axes data rates ($^\circ$/sec) which must be integrated over time to obtain the current heading, namely the *pitch-yaw-roll* values (cfr. Section 5.2). Axis a heading at time $\bar{t}$ is therefore obtained as

$$\eta_a(\bar{t}) = \sum_{t=1}^{\bar{t}} \varrho_a(t) \times \Delta T \tag{6.1}$$

where η_a, ϱ_a are respectively axis a heading and angular rate, $t = 1, \ldots, \bar{t} - 1$ are the time instants prior to the current one and ΔT is the integration period, which in our case is $T = 0.0116$s. This solution allows us to have inherently syncronized audio and orientation tracking signals, with an orientation information every 512 audio samples. Note that the gyroscope has a significant *drift* (time-dependent increasing error), which needs to be compensated. Due to the fact that this drift is not perfectly linear and might change depending on various factors, before starting the acquisition the gyroscope is left in rest position for a a period of time (about $3-5$ minutes) in order to obtain a reference drift value. We noticed however that this drift actually constitutes an issue. Its variation over time is not regular and might in some cases even change slope. Within the context of a violin perfomance, evaluating the impact of this drift on the overall orientation tracking is nearly impossible. The acquisition aimed at estimating the drift is a first starting point, but it proved to be not enough in many situations. A better solution is to take into account also the information provided by the compass (available on the same device) and integrate it

to the gyroscope data: this allows a more reliable estimation on the whole.

Once the acquisition is completed, we therefore have three files: one for the audio signals, one for the orientation trackings and one for the drift measurement.

These files are processed by a *Matlab* script, that performs all the steps presented in Chapter 5: frames selection, source localization and orientation estimation, energy measurement, distance and dynamic compensation. The data is then visualized through a series of plots, as we show in the next Sections.

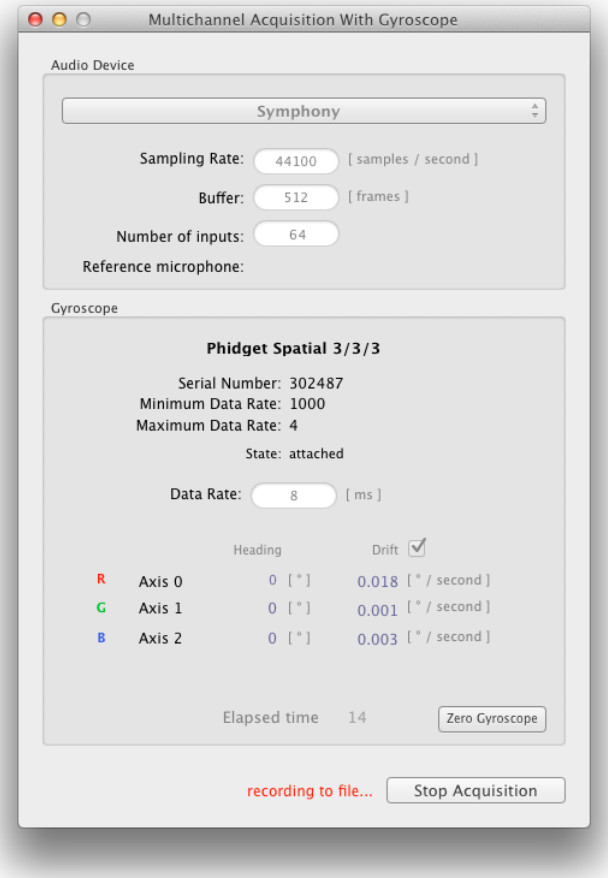


Figure 6.9. Acquisition application: graphical interface.

6.3 Results

In this section we present the results of the experiment that we made to validate our radiation pattern estimation method. We estimated the two dimensional radiation pattern of a *Genelec 8020A* loudspeaker (see Appendix B for specifications) at different frequencies. The technique we used is the one presented in Chapter 5: using multiple-camera plenacoustic methods, the system performs autocalibration, 3D localization and pattern estimation. The setup is the one presented in Section

6.2, with the second camera centered in $[-0.75\text{ m}, 1.2\text{ m}]^T$ and the angle between the two cameras equal to 90 °.

The experiment was performed into a semi-anecoic room, where walls are covered in absorbing acoustic foam, while floor and ceiling are to be considered reflectors. This fact influenced the results, as the vertical reflections introduced a noise that cannot be considered negligible.

6.3.1 Autocalibration

Autocalibration was performed through a dedicated acquisition. We positioned the loudspeaker so that the mid-range cone was at the same height of the third array from the top of both cameras (1.53 m). Assuming that the cameras were perfectly vertical and at the same height from the ground, there is no loss of generality in taking into account only one ULA per camera to perform autocalibration, as only the x, y translation and rotation need computing. We therefore applied the method presented in Section 4.3, having a white noise emitted by the loudspeaker in 13 different positions in front of the cameras. The resulting calibration errors are:

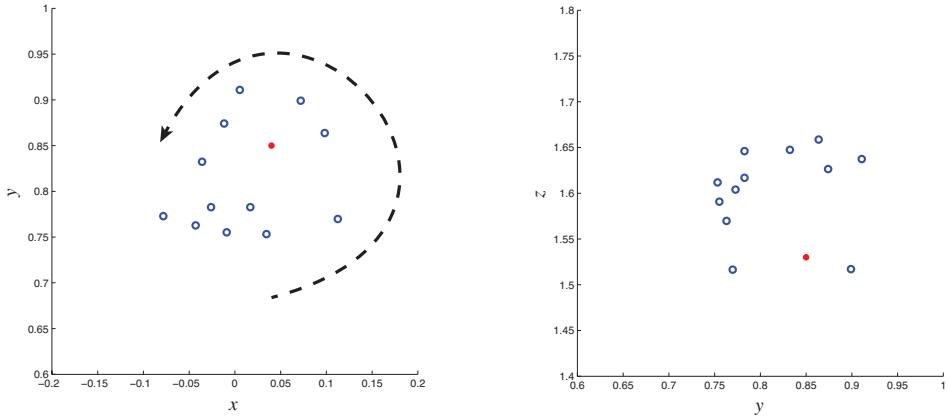
$$\begin{aligned} \text{Orientation estimation error} &: 2.02\ ^\circ \\ \text{Center estimation error} &: 0.05\text{ m}. \end{aligned} \tag{6.2}$$

Note that this errors are computed with respect to a manual measurement of the setup that might not be perfectly accurate itself. Furthermore, the two cameras might have been not perfectly vertical: due to the conformation of the tripods used to hold the cameras, a precise positioning was not possible.

6.3.2 Loudspeaker Position Estimation and Radiation Pattern Measurement

The loudspeaker was positioned at $[0.04\text{ m}, 0.85\text{ m}, 1.53\text{ m}]^T$, with the gyroscope fixed on top of it. In an iterative fashion, the louspeaker emitted a white noise signal of fixed length (5 s) and was rotated manually on its vertical axis of about 10 ° counter-clockwise. As a first step, we estimated the 3D position of the louspeaker for each acquisition. The results of the localizations are shown in Figure 6.10 (note that the reference camera is here considered to lie on the x axis). The dashed line represents the sense or the rotation. We notice that the estimated positions follow the mid-range cone of the loudspeaker as it rotates. The distance between opposite localizations (e.g. top-left and bottom-right) is in fact approximately the width of the chassis (10 cm). In general the localizations are consistent with the manual rotations that we made. The localizations in the lower-left sector are also consistent, as the last few rotations were smaller. The points on the left are more distant from the initial position as a consequence of the loudspeaker tripod oscillation on the vertical axis. As for the vertical estimation (b), this is also consistent with the conformation of the loudspeaker: the distance between mid-range cone and tweeter cone is approximately 20 cm. This means that the source is overall localized in the middle between the two components. The residual localization error is to be ascribed to the noise contribution of the vertical reflections of the room.

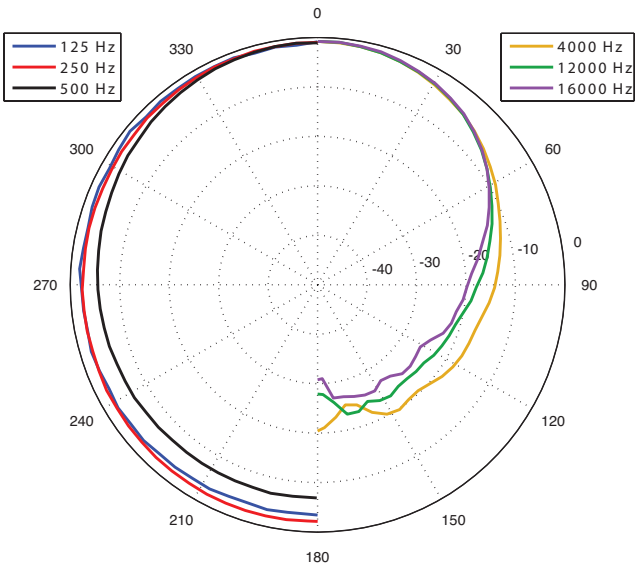
We then proceeded with evaluating the radiation pattern at six meaningful



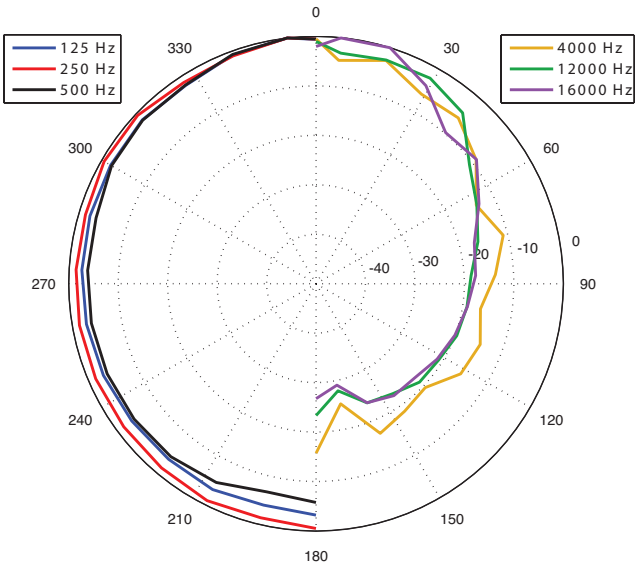
(a) *(b)*

Figure 6.10. Estimated loudspeaker 3D positions in the (x, y) and (y, z) planes. The red dot is the actual position of the mid-range cone.

frequencies: 125 Hz, 250 Hz, 500 Hz, 4K Hz, 12K Hz, 16K Hz. Figure 6.11 shows the results of the experiment, compared to nominal radiation patterns provided by the manufaturer, presented in polar diagrams. Note that we only display half of the pattern: this is due to the fact that it is simmetric both in the specifications and in our results. First, we focus on the low frequency patterns: we can observe that the quasi-omnidirectionality is preserved on all three cases and that the ratio between the three pattern is approximately the same as in the nominal patterns. In particular, in the rear part (180 °) the energy values correspond with a certain precision. The nominal high frequency patterns have a more complex shape and are again similar to each other. We can see that also in this case the results of our experiment overall coincide with the ones provided by the manufacturer. The irregular lines in the results are to be ascribed mainly to low resolution: the manual rotation of the loudspeaker was not precise and in some cases possibily too wide. Even so, the result shows that the estimate pattern is in general very accurate.



(a) Nominal



(b) Experimental Results

Figure 6.11. Loudspeaker experimental results (b) versus the spcifications provided by the manufacturer (a)

Chapter 7

Conclusions

Plenacoustic imaging is a powerful tool. It is capable of condensing the vast information about the soundfield into a compact and immediate representation that can be used at many different levels. In this thesis we have seen a meaningful example on how the one-shot acquisition of a set of plenacoustic images allows us to perform a series of different analyses on the acoustic scene; we are able to detect and localize a sound source in 3D space by combining the 2D localizations of many linear arrays, and we are also able to estimate the energy that this source radiates towards those arrays by simply evaluating the peaks on the same plenacoustic images used for the localization. Of course this is possible thanks to the multiple-camera approach, that, by providing multiple views of the acoustic scene, improves the effectiveness of the analysis tools that were available so far. The main improvement is in source localization: single camera-based localization techniques can estimate DOA accurately, but they do not attain very good performances in distance estimation as the source gets farther away from the observation window. When adding more than one point of view on the acoustic scene, the joint information on DOA compensates significantly the distance estimation error, rendering the overall localization far more accurate. After the study and first validation through a series of simulations, we determined that the multiple-camera localization technique is applicable to a radiation pattern estimation system.

The extension to a multiple-camera scenario and the integration of the new plenacoustic techniques, allow the system to be more coherent and self-contained: the previous hybrid paradigm, where soundfield analysis relies on information retrieved through a depth map camera, is now replaced by a fully plenacoustic paradigm that exploits the ray-space representation of the soundfield in all the autocalibration, localization and pattern estimation steps. We want to stress that the application of the concepts introduced in this work to the pattern estimation system has served as a demonstration on how the extension to the multiple-view scenario widens the scope of plenacoustic imaging, rendering the soundfield analysis techniques more and more versatile. With our thesis, we laid a solid foundation for future works in the field of plenacoustic analysis.

Future Works

A very relevant subject that needs further investigation is the extension to three dimensions of the plenacoustic concepts introduced in this thesis, as well as, in general, of the ones already present in the literature. While our approach to 3D source localization is proven valid, it still is an hybrid solution that certainly does not completely exploit the potentialitis of plenacoustic analysis. A fully plenacoustic approach, which was out ot the scope of this thesis, would be a useful subject for future studies.

As far as plenacoustic radiation pattern estimation is concerned, it needs to be extended to the context of a real musical performance. Again, this aspect was out of the scope of our studies, and it certainly is the next step to be undertaken in an applicative work. Another future development could be taking plenacoustic radiation pattern estimation a step further: so far only the farfield radiation has been evaluated, considering the instrument as a monopole. This, as explained in this thesis, is a valid assumption in the *Fraunhofer* region, and enables the definition of the overall interaction of the instrument with the sourrounding environment. If on the other hand the plenacoustic camera was brought to the nearfield region, it would be possible to evaluate the energy radiate by each component of the instrument. Considering, as an example, the sound board of a violin, this kind of analysis could even be able to determine its main normal modes of vibration.

Appendix A

Spatial Signal Processing

For completeness' sake we now want to elaborate the concept of pseudospectrum that we tuched upon throught the thesis and in particular in the midst of section 3.2. Furthermore we introduce a general overview of the principles of spatial signal processing. We here take as reference one of the ULA sub-arrays and therefore drop the index w relative to its position in the main array. Also, for simplicity's sake, we'll assume that the first microphone instead of the middle one is considered as the reference. Let $x(t)$ denote the value of the signal waveform sampled at the reference point at time t and τ_n the time needed for the wave to travel from the reference point to microphone j ($j = 1, \ldots, J$). Considering microphone j and neglecting its impulse response, its output can be written as

$$s_j(t) = x(t - \tau_j) + e_j(t) \tag{A.1}$$

where $e_j(t)$ is an additive error. The kth frequency component of the signal is

$$s_j(t, \omega_k) = e^{-j[\omega_k(t-\tau_j)+\phi]} + e_j(t, \omega_k) = e^{-j\omega_k\tau_j}x(t) + e_j(t, \omega_k) \tag{A.2}$$

where $e^{-j\omega_k\tau_j}$ accounts for propagation delay from the reference point to the nth microphone. $e^{-j\omega_k\tau_j}$ evidently embeds information about the position of the source. We collect the propagation terms of all microphones into a propagation vector

$$\mathbf{a}(\theta, \omega_k) = \left(e^{-j\omega_k\tau_1}, \ldots, e^{-j\omega_k\tau_W}\right)^T \tag{A.3}$$

where the variable θ is the Direction Of Arrival (DOA) of the source. Now, let d denote the the distance between two consecutive microphones and θ the DOA of the signal illuminating the array, under the planar wave hypotesis we find that

$$\tau_j = (j-1)\frac{d\sin(\theta)}{c} \quad , \quad -\pi/2 \leq \theta \leq \pi/2, \tag{A.4}$$

where c is the sound speed. The propagation vector can then be rewritten as

$$\mathbf{a}(\theta, \omega_k) = \left(1, e^{-j\omega_k d\sin(\theta)/c}, \ldots, e^{-j(J-1)\omega_k d\sin(\theta)/c}\right)^T. \tag{A.5}$$

Furthermore, if $\lambda = c/f_k$ is the wavelength and if we define the quantity ω_{k_s} as

$$\omega_{k_s} = \omega_k \frac{d\sin(\theta)}{c}, \tag{A.6}$$

we obtain yet another form of the propagation vector:

$$\mathbf{a}(\theta) = \left(1, e^{-j\omega_{k_s}}, \ldots, e^{-j(J-1)\omega_{k_s}}\right)^T \tag{A.7}$$

which is completely analogous to the vector built sampling the sinusoidal signal $e^{-j\omega_k t}$. For this analogy, the quantity ω_{k_s} is referred to as *spatial frequency*. If we want to sample a continuous-time sinusoidal signal with frequency f_k, then, in order to avoid aliasing effects, the sampling frequency F_s should satisfy the condition $F_s \geq 2f_k$. Also in the spatial sampling a problem of minimum sampling frequency arises. We have to guarantee that

$$|f_{k_s s}| \leq 1/2 \iff d\sin(\theta) \leq \lambda/2, \tag{A.8}$$

which is always satisfied if

$$d \leq \lambda/2. \tag{A.9}$$

In other words, the ULA is perfoming a uniform spatial sampling of the wavefield and the distance between micrphones, which can be though of as the spatial sampling period, must be smaller than half the signal wavelength in order to avoid aliasing. In the light of the concept of spatial sampling, we can state that there is an analogy between temporal temporal and spatial filtering. In fact, in the same way that a temporal filter can be designed to enhance or attenuate selected frequencies, a spatial filter can be designed to enhance or attenuate the signal coming from selected DOA. Being $\mathbf{h}$ the set of filter weights, the spatially filtered output of an array illuminated by a narrowband wavefront with complex envelope $u(t)$ and DOA equal to θ is given by

$$s_f(t, \omega_k) = \left(\mathbf{h}^H(\omega_k)\mathbf{a}(\theta, \omega_k)\right)u(t, \omega_k) \tag{A.10}$$

which is the equation of the spatial filter, also called *beaformer*. The power $E[|s_f(t)|^2]$ of the spatially filtered signal should give a good indication of the energy coming from direction θ:

$$E[|s_f(t)|^2] = \mathbf{h}^H(\omega_k)\mathbf{R}(\omega_k)\mathbf{h}(\omega_k), \quad \mathbf{R}(\omega_k) = E[\mathbf{s}(t)\mathbf{s}^H(t)] \tag{A.11}$$

Where $\mathbf{R}(\omega_k)$ is the covariance matrix and $\mathbf{s}(t, \omega_k)$ the kth frequency component of the array's output vector

$$\left(s_1(t, \omega_k), \ldots, s_W(t, \omega_k)\right)^T. \tag{A.12}$$

$E[|s_f(t)|^2]$ should peak at the DOA(s) of the source(s) located in the array's viewing field when evaluated over the DOA range of interest.
We can state that a signal coming from a DOA θ passes undistorded if

$$\mathbf{h}^H(\omega_k)\mathbf{a}(\theta, \omega_k) = 1. \tag{A.13}$$

In order to minimize the energy coming from all the other DOAs, we can follow several approaches. We here present the two that are of interest for our purposes.

Delay-And-Sum (DAS) is a data independent method and has the objective of finding $\mathbf{h}(\omega_k)$ that minimizes $\mathbf{h}^H(\omega_k)\mathbf{h}(\omega_k)$, subject to $\mathbf{h}^H(\omega_k)\mathbf{a}(\theta,\omega_k) = 1$. Assuming that the steering vector has been normalized so that $a^H(\theta,\omega_k)a(\theta,\omega_k) = J$ we have $\mathbf{h}(\omega_k) = \mathbf{a}(\theta,\omega_k)/W$, which inserted in the power formula leads to

$$E[|s_f(t)|^2] = \frac{\mathbf{a}^H(\theta,\omega_k)\mathbf{R}(\theta,\omega_k)\mathbf{a}(\theta,\omega_k)}{J^2}. \tag{A.14}$$

Though we the theoretical covariance matrix cannot be exactly determined from the available finite sample $\mathbf{s}(t)_{t=1}^T$, therefore it must be replaced by the estimate

$$\hat{\mathbf{R}}(\omega_k) = \sum_{t=1}^{T} \mathbf{s}(t,\omega_k)\mathbf{s}^H(t,\omega_k). \tag{A.15}$$

Omitting J^2, we can write the DAS pseudospectrum function as:

$$\mathbf{h}_{DAS}(\theta,\omega_k) = \mathbf{a}^H(\theta,\omega_k)\hat{\mathbf{R}}(\omega_k)\mathbf{a}(\theta,\omega_k). \tag{A.16}$$

DAS is useful when the DOA is not accurately know since it has a wider main lobe. As an example it will be used for radiation pattern measurement, as the localization method is not able to provide a precise DOA estimate. On the contrary, for the plenacoustic image computation presented at the beginning of section 3.2 higher resolution is needed. The next method is therefore used.

The *Minimum Variance Distortionless Response* (MVDR) beaformer is a data-dependent method and it aims at minimizing the whole term $\mathbf{a}^H(\theta,\omega_k)\mathbf{R}(\omega_k)\mathbf{a}(\theta,\omega_k)$. The solution to this optimization problem can be written as

$$\mathbf{h}(\theta,\omega_k) = \frac{\mathbf{R}^{-1}(\omega_k)\mathbf{a}(\theta,\omega_k)}{\mathbf{a}^H(\theta,\omega_k)\mathbf{R}^{-1}(\theta,\omega_k)\mathbf{a}(\theta,\omega_k)} \tag{A.17}$$

which, given the same observations made for DAS, leads to the MVDR pseudospectrum formula:

$$\mathbf{h}_{MVDR}(\theta,\omega_k) = \frac{1}{\mathbf{a}^H(\theta,\omega_k)\hat{\mathbf{R}}^{-1}(\omega_k)\mathbf{a}(\theta,\omega_k)}. \tag{A.18}$$

Appendix B

Technical Specifications

Loudspeakers

Table B.1. Experimental setup: loudspeakers specifications.

Model	**Genelec 8020A**
Total Power Output	RMS 20 W × 2
Total Harmonic Distortion	≤ 0.08%
Signal-to-Noise ratio	≥ 95 dB
Bass Driver	105 mm cone
Treble Driver	19 mm metal dome

Microphones

Table B.2. Experimental setup: microphone *Beyerdynamic MM1* specifications.

Model	**Beyerdynamic MM1**
Transducer type:	condenser (back electret)
Polar pattern	omnidirectional
Frequency range:	20 Hz to 20 kHz
Sensitivity:	15 mV/Pa (−36.5 dBV)
Max SPL for 1% THD:	128 dB
Signal-to-Noise ratio (rel. to $1Pa$):	57 dB
Nominal impedance:	330 Ω
Powering:	12 V to 48 V phantom power
Connector:	3-pin XLR
Dimensions (diameter):	9 mm($head$)
Dimensions (length):	133 mm

Microphone Preamplifier

Table B.3. Experimental setup: microphone preamplifier *Aphex 188* specifications.

Model	**Aphex 188**
Gain:	26 dB to +65 dB
Input Impedance:	Variable
EIN:	−125 dBu
Frequency Response:	Not specified by manufacturer

A/D and D/A Converter

Table B.4. Experimental setup: *Symphony I/O* specifications (Analog I/0).

Model	**Symphony I/O**
	32 inputs and 32 outputs
Type:	Balanced through Apogee's proprietary Perfect Symmetry Circuitry
Level:	+4 dBu nominal / −24 dBu max. or −10 dBV nominal / +6 dBV max
Input Impedance:	10 kΩ
Output Impedance:	25 Ω

Table B.5. Experimental setup: *Symphony I/O* specifications (Analog to Digital).

Model	**Symphony I/O**
Frequency response:	1 Hz to 20 kHz, +0/ − 0.05 dB
Dynamic range:	120 dB (A-weighted)
THD+N @ +20 dBu	−113 dB (0.00024%)

Table B.6. Experimental setup: *Symphony I/O* specifications (Analog Out performance).

Model	**Symphony I/O**
Frequency response:	*dc* to 20 kHz, +0/ − 0.05 dB
Dynamic range:	129 dB (A-weighted)
THD+N @ 20 dBu	−117 dB (0.00014%)

Table B.7. Experimental setup: *Symphony I/O* specifications (Connections).

Model	**Symphony I/O**
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 Ω BNC word clock input and output
MIDI	1 In and 1 Out, 5-pin female DIN connectors

Gyroscope

Table B.8. Experimental setup: gyroscope specifications.

Model	**Phidgets Spatial** 3/3/31044
Max Speed (X-axis,Y-axis)	±400 °/s
Max Speed (Z-axis)	±300 °/s
Resolution (X-axis,Y-axis)	0.02 °/s
Resolution (Z-axis)	0.013 °/s
White Minimum Drift	0.0042 °/s

Bibliography

[1] Thibaut Ajdler *The Plenacoustic Function and its Applications*

[2] Thibaut Ajdler, Martin Vetterli *The Plenacoustic Function, Sampling and Reconstruction*, Ecole Polythecnique Federale de Lousanne, facoulte informatique et communication, 2006

[3] G. Bissinger, M. Bailey *Testing a modal-analysis-based model to predict averaged violin radiation from mechanical response measurements*, 1997

[4] G. Bissinger, J. Keiffer *Radiation damping, efficency and directivity for violin normal modes below 4 khz*, Acoustical Society of America, 2002

[5] Stephen P. Boyd *Introduction to Linear Dynamical Systems - Course Reader*

[6] A. Perez Carrillo, J. Bonada, J. Patyen, V. Valimaki *Method for measuring violin sound radiation based on bowed glissandi and its appliocation to sound synthesis*, The Journal of the Acoustical Society of America, 2011

[7] D. F. Comesana, T. Takeuchi *Measuring musical instruments directivity patterns with scanning techniques*, ICSV19, Vilnius, LT, 2012

[8] Marco Compagnoni, Paolo Bestagini, Fabio Antonacci, Augusto Sarti, Stefano Tubaro *Localization of Acoustic Sources Through the Fitting of Propagation Cones Using Multiple Independent Arrays*, IEEE Transactions on audio, speech, and language processing, 2012

[9] H. Dunnwald *Deduction of objective quality parameters on old and new violins*, Catgut Acoust. Soc. J, 1991

[10] Richard Hartley, Andew Zisserman *Multiple View Geometry in computer vision*, Cambridge University Press, 2000

[11] Yiteng Huan, Jacob Benesty *Audio Signal Processing For The Next-Generation Communication Sytems*, Kluwer Academic Publishers, 2004

[12] Dejan Marcovic, Fabio Antonacci, Augusto Sarti, Stefano Tubaro *Soundfield Imaging in the Ray Space* 2012

[13] Luca Mucci *Estimation of the Radiation Pattern of a Violin during the Performance using Plenacoustic Methods*, Politecnico di Milano 2013

[14] Z. Otcencasek, V. Syrovy *Directivity of violin radiation*,

[15] Giorgio Sandrini *Acoustic Imaging in the RaySpace: Application to Environment Inference*, Politecnico di Milano, 2012

[16] Salvo Daniele Valente, Marco Tagliasacchi, Fabio Antonacci, Paolo Bestagini, Augusto Sarti, Stefano Tubaro *Geometric calibration of distributed microphone arrays from acoustic source correspondences*, 2010

[17] L. Wang, C. Burroghs *Acoustic Radiation From Bowed Violins*, Achitectural Engineering Faculty Publications, University of Nebraska, Lincoln, 2001

[18] G. Weinreich, E. B Arnold *Method for measuring acoustic radiation fields*, 1980

[19] G. Weinrich *Directional Tone Color*, 1997

[20] E. G. Williams *Fourier Acoustic - Sound Radiation And Nearfield Acoustic Holography*, Academic Press, 1999