Politecnico di Milano
University campus of Como

Master of Science in Computer Engineering

School of Industrial and Information Engineering

A Plenacoustic Approach to Sound Scene Manipulation

Master Graduation Thesis of:
Francesco Picetti
Candidate Id: 855153

Supervisor:
Prof. Fabio Antonacci

Assistant Supervisor:
Dr. Eng. Federico Borra

Academic Year 2017/2018

Politecnico di Milano
Polo territoriale di Como

Corso di Laurea Specialistica in Ingegneria Informatica

Scuola di Ingegneria Industriale e dell'Informazione

Un Approccio Plenacustico alla Manipolazione di Scene Sonore

Tesi di Laurea Specialistica di::
Francesco Picetti
Matricola: 855153

Relatore:
Prof. Fabio Antonacci

Correlatore:
Dott. Ing. Federico Borra

Anno Accademico 2017/2018

Abstract

Sound field manipulation is a research topic that aims at modifying the properties of an acoustic scene. In the literature it has been proven that physically accurate sound field representations become unpractical when complex sound scenes are considered, due to the enormous computational effort they would require.

This thesis aims at developing a novel manipulation methodology based upon *Soundfield Imaging* techniques, which model the acoustic propagation adopting the concept of acoustic rays, i.e. straight lines that carry acoustic information. We have taken advantage of the *Ray Space Transform*, a powerful tool that efficiently maps the sound field information acquired by means of a Uniform Linear Array (ULA) of microphones onto a domain known as ray space. Points in this domain represent rays in the geometric space, therefore the analysis of the propagating sound field could be addressed considering each directional component individually.

The description of acoustic scenes is here approached adopting a parametric spatial sound paradigm. More precisely, the sound scene is described by defining the acoustic source, its position and orientation in space, its emitted signal and finally its radiation pattern, defined as the angular-frequency dependence of the amplitude of the signal.

First, the ray space transform of the microphone signals is computed and provided to the parameter estimation stage. In particular, the source position is estimated through a weighted least squares problem, while the radiation pattern is extrapolated from the linear pattern of the ray space image. Since the microphone array "sees" the source from a narrow range of directions, the radiation pattern is modelled in the *Circular Harmonics* domain and the coefficients are estimated through an optimization problem. Finally, a spatial filter called *beamformer* is designed upon the results of the previous steps in order to extract the source signal directly from the array signals.

The estimated paramaters could be provided to any parametric sound scene rendering system. Through computer-aided simulations and experiments we have proven the capability of the proposed system performing different manipulations. The results of the tests suggest that the proposed methodology would play a significant role in future research topics on spatial audio processing.

Sommario

La manipolazione dei campi acustici è un argomento di ricerca che si prefigge di modificare le proprietà di una scena acustica. In letteratura è stato comprovato che le rappresentazioni fisicamente accurate del campo sonoro risultano impraticabili nella modellazione di scene complesse, a causa dell'ingente costo computazionale richiesto.

Questa tesi sviluppa una metodologia di manipolazione innovativa basata sulle tecniche di *Soundfield Imaging*, nelle quali la propagazione sonora sfrutta il concetto di raggi acustici, ovvero linee rette che trasportano informazione acustica. Abbiamo adottato la trasformata *Ray Space*, un potente operatore che mappa l'informazione acquisita da una schiera lineare uniforme di microfoni nello *spazio dei raggi*. Punti in questo dominio rappresentano raggi nello spazio geometrico, perciò l'analisi della propagazione del campo acustico può essere affrontata considerando individualmente le componenti direzionali.

La descrizione della scena acustica fa proprio il paradigma parametrico della spazializzazione sonora. Più precisamente, la scena acustica è qui parametrizzata definendo la sorgente acustica, la sua posizione e il suo orientamento nello spazio, il suo segnale emesso e infine il *pattern* di radianza, definito come la dipendenza dell'ampiezza del segnale dalla frequenza e dalla direzione di propagazione.

Innanzitutto i segnali microfonici viengono trasformati e l'immagine ray space processata per stimare i parametri. In particolare, la posizione della sorgente è stimata grazie a una regressione ponderata dei minimi quadrati, mentre il pattern di radianza è estrapolato dalla linea emergente dall'immagine ray space. Siccome la schiera di microfoni "vede" la sorgente in un intervallo ristretto di direzioni, il pattern di radianza è modellato nel dominio delle *Armoniche Circolari* e i coefficienti stimati con un problema di ottimizzazione. Infine, un filtro spaziale chiamato *beamformer* viene progettato conoscendo i risultati dei passaggi precedenti al fine di estrarre il segnale della sorgente direttamente dai segnali microfonici.

I parametri stimati possono essere utilizzati da qualsiasi sistema parametrico di *rendering* acustico. Grazie a simulazioni computerizzate ed esperimenti in laboratorio abbiamo dimostrato le capacità del sistema proposto eseguendo diverse manipolazioni. I risultati delle prove suggeriscono che la metodologia proposta giocherà un ruolo importante nelle ricerche future sull'elaborazione di campi acustici.

Ringraziamenti

Questa tesi è stata sviluppata presso l'*Image and Sound Processing Lab* del Politecnico di Milano ed è il risultato di un lungo percorso accademico. Vorrei innanzitutto ringraziare il mio relatore, prof. Fabio Antonacci, per avermi dato l'opportunità di lavorare ad un argomento così interessante. Un rigraziamento speciale merita il mio correlatore, Federico, per l'inestimabile pazienza e il prezioso supporto offertomi nei momenti più difficili di questo lavoro.

La mia gratitudine va anche ai miei compagni di corso e a tutto il gruppo ISPL-ANTLAB. Ricorderò particolarmente Clara, Stella, Jacopo e Luca; grazie per gli incoraggiamenti e l'affetto, per le discussioni sui grandi sistemi e sui piccoli sintetizzatori.

Grazie alla mia famiglia, ai miei genitori e ai miei fratelli, per l'immenso supporto e sostegno. Non avrei raggiunto questo traguardo senza di voi. Infine una dedica speciale va ad Alice, che rende la vita migliore.

F.P.

Dimenticavo. Grazie a Wolfgang che dà un senso a tutto questo.
Anche se è morto.

Contents

List of Figures

List of Tables

Introduction

Processing of sound field is a research topic that aims at manipulating the sound field in order to extract information about the overall acoustic characteristics of both the environment and the objects in a scene. The ability to sense spatial properties of sound enables humans to experience immersivity, i.e the sense of presence in a scene. In the last decades this topic has grown interest in both research and industry communities, since it can be applied to a great variety of markets and purposes. As an example, in multimedia and entertainment industry we would cite live music venues, movie theatres, videogames, and teleconference. Moreover, this topic is experiencing a growth in interest also in sound and vibration control, e.g. in the automotive industry.

This thesis addresses the problem of manipulating a sound scene and proposes a novel synthesis-by-analysis methodology based on geometrical acoustics. In order to capture the spatial cues of the sound propagation, a suitable field representation is needed. Among all the acoustic field representations, we can distinguish between two main paradimgs. The so called nonparametric representations describe the acoustic field as a function of space and time and assume no a-priori knowledge on the acoustic scene. These physically-motivated reprepresentations decompose the sound field on different basis functions. We would cite *Ambisonics* and its followers, which decompose the sound field in terms of spherical harmonics, and *WaveField Synthesis* that is based on the Huygens' principle, which states that the wavefront of a primary source can be reconstructed through a conveniently-driven spatial distribution of secondary sources, e.g. loudspeakers. Although their physical motivation and precision, the nonparametric representations require an enormous computational effort in order to solve exactly the wave equation and its boundary conditions, especially in modelling complex acoustic environments.

The second paradigm is the geometric representation and it models the acoustic propagation upon the concept of *acoustic rays*, i.e. straight lines that carry acoustic information. Although this view provides a rough approximation of an acoustic field, in literature it has been demonstrated that this approach can model complex acoustic environments, for which a physically accurate modelling is not practical. This is the aim of *Plenacoustic Imaging* [2], a novel technique that maps plane-wave components of the acoustic field to points in the *ray space*, defined as

a projective domain whose primitives are acoustic rays. If we want to measure the *plenacoustic function* in a given point, we can do so by using a microphone array centred in that point, and estimating the radiance along all visible directions through spatial filtering. By subdividing the array into small sub-arrays the plenacoustic function can be acquired over every point belonging to the microphone array boundaries. It has been proven [3] that the main acoustic primitives (such as sources and reflectors) appear in the ray space as linear patterns, thus enabling the use of fast techniques from computational geometry to deal with a wide range of problems, such as localization of multiple sources and environment inference. One step beyond, in [4] the authors devise a linear operator called *Ray Space Transform* that maps the signals acquired from a Uniform Linear Array (ULA) of microphones in the ray space domain. This transformation is based on a short space-time Fourier transform using Gabor frames, enabling the definition of analysis and synthesis operators that exhibit perfect reconstruction capabilities. In other words, the array is subdivided into shifted and modulated replica of a prototype spatial window. Then, for each sub-array, a spatial filter called *beamformer* scans all the visible directions and encodes the sound field. For the purpose of this thesis, the geometrical acoustics can give us a incisive advantage, because by capturing each ray in a single position in space, the acoustic scenario can be easily reconstructed also in near-field conditions.

In its simplest formulation, a sound scene is described as a source emitting acoustic signals in space. Through a ULA of microphones, the analysis stage aims at localizing the source in space and extracting its signal. Figure 1 help visualizing the scene and the transformation we aim at applying to it. The manipulation can be described as a three-steps process. First, the microphone array signal is transformed in the ray space; the localization process can be easily performed trough a weighted least squares regression on the peaks that emerge from the ray space image. Second, the ray space image is read along the linear pattern that is dual of the source position in the geometric space. However, we cannot assume the source signal propagation to be constant over the directions. Due to the physical characteristics of the source, the sound propagation prefers some directions among others. Morever, these inequalities in sound propagation depends on the time frequency; indeed, the acoustic absorption of the source body depends on the frequency. In order to model these phenomena we adopt the concept of *Radiation Pattern* as an angular-frequency dependence of the amplitude of the signal emitted (or received) by any acoustic object. The radiation pattern can also describe the orientation of the source in the space; by rotating the radiation pattern we can rotate the source about its axes. The ray space components extracted in the second step is thus interpreted as the source *pattern-times-signal*. In order to extrapolate the radiation pattern weights that mask the pure signal, we adopt the *Circular Harmonics Decomposition*. The extracted ray space components are compensated for the acoustic

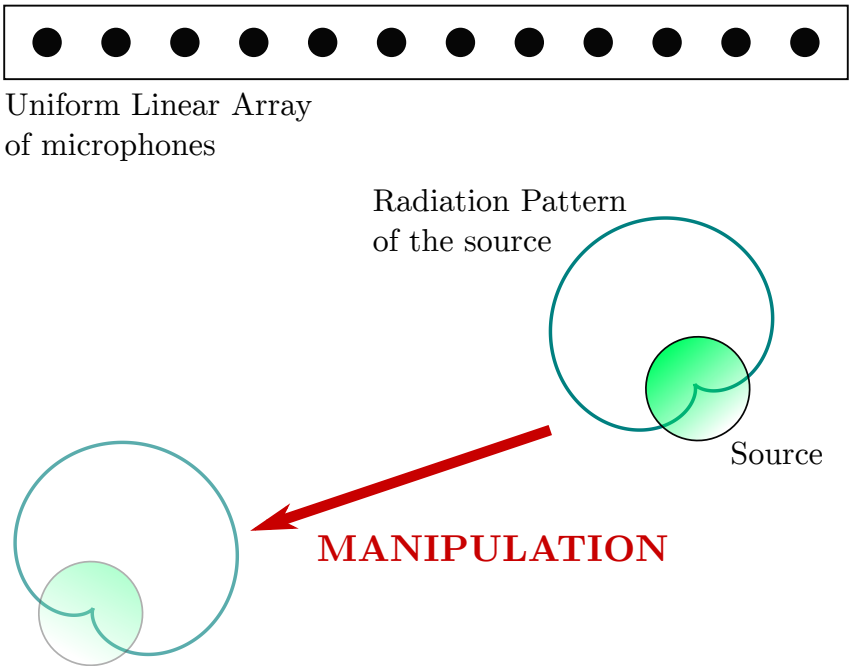


Figure 1: An example of acoustic scene and manipulation.

channel and then fed to a linearly constrained quadratic programming problem that estimates a suitable set of coefficients in the Circular Harmonics domain. This way, the radiation pattern of the source seen by the microphone array is reconstructed by a simple matrix multiplication. The third step takes advantage of both the localization and pattern extrapolation procedures. In particular, the source signal is extracted from the array signal by a simple beamformer that exploits the acoustic transfer function from the source to the microphones position, compensated for the radiation pattern attenuations in the set of directions from the source towards the array.

The sound scene analysis results in the estimate of the position, radiation pattern and signal of the acoustic source. These parameters can be easily modified and fed to any parametric sound scene rendering system. In fact, the sound field can be recostructed in any point of the region of interest of the ULA. In order to evaluate the validity of the manipulation, we sample the sound field with the same microphone array of the anylisis stage; more precisely, we compare the desired sound field with the synthesized version of the same acoustic scene. The extracted signal is propagated from the desired source position to the microphone array; the synthesized microphones are then weighted for the radiation pattern visible by the array and desiderably oriented. The desired radiation pattern weights are computed in the Circular Harmonics domain using the coefficients inherited from the analysis stage.

Through computer-aided simulations and experiments, it emerges that the chosen representation allows the manipulation of the sound scene in a very intuitive fashion. With respect to the state of the art, this work

extends the plenacustic processing of sound fields extrapolating the radiation pattern directly from the Ray Space Transform of the microphone array signal, opening to new possibilities of processing the sound scene directly in the ray space. Moreover, the parametric spatial sound paradigm adopted in this thesis is proven to be reliable and intuitive.

This thesis is organized as follows. In Chapter 1 we provide an overview on the nonparametric sound field representation as the theoretical foundations of the sound field analysis and synthesis techniques present in literature. Then, moving to the realm of geometrical acoustics, we introduce the reader to the plenacoustic imaging techniques that is the solid root of the methodology devised in this thesis. Chapter 2 addresses the theoretical background of the ray space transform, along with an example of application; moreover, we briefly describe the theory and motivations of the acoustic radiation pattern. Chapter 3 is entirely dedicated to the description of the developed manipulation methodology. More precisely, the chapter starts with a description of the parametric spatial sound paradigm useful for describing any acoustic scene; then the analysis and synthesis steps are presented. In Chapter 4 we provide some simulations and experiments designed for proving the proposed system capabilities. More precisely, we show the manipulation behavior when the analysed sound scene is rotated and translated during the synthesis stage. Finally in Chapter 5 we draw the conclusions and outline possible future research directions.

1
State of the Art

This chapter introduces the models suitable for describing the generation and propagation of sound. In particular, we describe in detail the theoretical foundations of the techniques developed so far for the purpose of this work, i.e. the manipulation of sound scene. The first part reviews the wave equations and their simplest solutions (the plane wave and the Green function) along their characteristics. The second section reviews the two sound reproduction techniques that have emerged in the literature from the '70s, *Ambisonics* and, later, *WaveField Synthesis*. In order to overcome the strict limitations of the classical stereophony, these physically-motivated techniques assumes the wavefield to be a superposition of elementary components, in particular Ambisonics decomposes the soundfield through spherical harmonic functions [5], while WaveField Synthesis exploits the Huygens principle, which states that any wavefront can be decomposed into a superposition of elementary spherical wavefronts emitted by secondary sources [6].

The last sections moves from the realm of wave acoustics to the realm of the geometric acoustics, which introduces a novel signal representation paradigm based on the concept of acoustic ray and briefly describes the advent of the plenacoustic imaging.

In this chapter the sound is described by a scalar function, the *acoustic field*, whose domain is the union of time and space regions:

$$p(\mathbf{r}, t), \quad \mathbf{r} \in \mathbb{R}^3, \quad t \in \mathbb{R} \tag{1.1}$$

is the general form of an acoustic field, with $\mathbf{r}$ and t denoting the spatial coordinates and time, respectively. Since the scenario considered in this work is that of a sound field restricted to a plane, we consider only two spatial coordinates that in Cartesian and polar coordinate systems

are denoted by

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} \rho \\ \theta \end{pmatrix} \tag{1.2}$$

The polar coordinates are related to Cartesian coordinates by the following relationships:

$$\mathbf{x} = \rho \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad \mathbf{r} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \arctan\left(y/x\right) \end{pmatrix} \tag{1.3}$$

In the linear regime, i.e. the medium of propagation behaves linearly, the acoustic field can be described by small-amplitude variations of the pressure; considering a volume without any source, in order to be a valid acoustic field in equation (1.1) must satisfy the *homogeneous wave equation*

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0 \tag{1.4}$$

where c denotes the speed of propagation in ms^{-1} and ∇ is the Laplace operator. In many context it is useful to consider acoustic propagation as a function of space and temporal frequency ω. The sound fields are denoted by $p(\mathbf{r}, t) = P(\mathbf{r}, \omega)e^{j\omega t}$, being $P(\mathbf{r}, \omega) = \mathcal{F}_t \{p(\mathbf{r}, t)\}$ the Fourier Transform with respect to the time variable t. The correspondent equation to the (1.4) in the transformed domain can be obtained as:

$$\mathcal{F}_t \left\{ \nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} \right\} = 0 \tag{1.5}$$

eploiting the linearity property and the differentiation property of the Fourier Transform and substituting the definition of $P(\mathbf{r}, \omega)$ it results

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = 0. \tag{1.6}$$

The equation above is known as *Helmholtz equation*.

If the volume under consideration is not free of sources, the wave equation loses its validity. In order to include a source excitation term, the right-hand side of (1.4) is replaced with a suitable source term depending on the volume flow rate per unit volume $q(\mathbf{r}, t)$, yielding the *inhomogeneous wave equation*

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -\frac{\partial q(\mathbf{r}, t)}{\partial t} \tag{1.7}$$

and its related *inhomogeneous Helmholtz equation*

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = -j\omega Q(\mathbf{r}, \omega). \tag{1.8}$$

1.1 Acoustic Field Representation

In this section we discuss some mathematical and physical solutions to equation (1.6) in the Cartesian reference frame, without loss of generality, and their main characteristics for analysis and synthesis tasks.

1.1.1 Plane Waves Decomposition

A simple canditate solution to the equation (1.6) is in the form of complex exponential function

$$P(\mathbf{r},\omega) = e^{j\langle \mathbf{k},\mathbf{r}\rangle}, \tag{1.9}$$

where $\mathbf{k}$ is called the *wavenumber vector* and the unit vector $\widehat{\mathbf{k}} = \mathbf{k}/\|\mathbf{k}\|$ is identified as the Direction of Arrival (DoA) of the plane wave. For sound fields in the form expressed by equation (1.9) to be solutions of the Helmholtz equation, $\mathbf{k}$ must be in $\mathbb{R}^2$ and satisfy the dispersion relation:

$$\|\mathbf{k}\|^2 = \left(\frac{\omega}{c}\right)^2. \tag{1.10}$$

Upon varying $\mathbf{k}$ in $\mathbb{R}^3$ constrained by the dispersion relation, one obtains a complete set of basis functions over which an arbitrary acoustic fields may be decomposed. The wave equation (expressed with respect to either time or temporal frequency) is linear, therefore general solutions can be obtained as superposition of plane waves of different frequencies and directions:

$$P(\mathbf{r},\omega) = \left(\frac{1}{2\pi}\right)^3 \iiint_{\mathcal{D}} \tilde{P}(\mathbf{k}) e^{j\langle \mathbf{k},\mathbf{r}\rangle} d^3\mathbf{r}, \quad \mathcal{D} = \left\{\mathbf{k}\in\mathbb{R}^3 : \|\mathbf{k}\| = \frac{\omega}{c}\right\}. \tag{1.11}$$

In the literature, the expansion in (1.11) is also known as Whittaker's representation.

1.1.2 Green Function

Arbitrary solutions to the inhomogeneous Helmholtz equation (1.8) are constructed starting from a basic solution, resulting by imposing a spatial impulse at location $\mathbf{r}'$ as the excitation term, i.e.

$$Q(\mathbf{r},\omega) = \delta(\mathbf{r}-\mathbf{r}')\delta(\omega),$$

where $\delta(\cdot)$ is the Dirac delta function. The sound field $G(\mathbf{r}|\mathbf{r}',\omega)$ is solution to the inhomogeneous Helmholtz equation obtained upon substituting the spatial impulse into (1.8):

$$\nabla^2 G(\mathbf{r}|\mathbf{r}',\omega) + \left(\frac{\omega}{c}\right)^2 G(\mathbf{r}|\mathbf{r}',\omega) = -\delta(\mathbf{r}-\mathbf{r}'). \tag{1.12}$$

The function $G(\cdot)$ is called *Green function*; under free space conditions it takes the form:

$$G(\mathbf{r}|\mathbf{r}',\omega) = \frac{e^{-j(\omega/c)\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|} \tag{1.13}$$

As for the homogeneous case, general solutions can be obtained by superposition of acoustic monopoles, described by different Green functions.

Finally, we describes the directional gradient of $G(\mathbf{r}|\mathbf{r}',\omega)$, as an example, with respect to r [6]:

$$\frac{\partial G(\mathbf{r}|\mathbf{r}',\omega)}{\partial r} = \frac{1}{4\pi}\left(j\frac{\omega}{c} - \frac{r}{\rho}\right)\frac{e^{-j\omega/c\rho}}{\rho^2}, \tag{1.14}$$

ρ being the distance between $\mathbf{r}$ and $\mathbf{r}'$, thus $\rho = \|\mathbf{r}-\mathbf{r}'\|$. Equation (1.14) can be interpreted as the spatial transfer function of a dipole source whose main axis is along the r-axis [7].

1.1.3 Kirchhoff-Helmholtz Integral

So far we have considered only the free-field scenario, i.e. where no sound reflections occur. However, there are situations in which one has to consider also the presence of acoustic boundaries. Two of the most important boundary conditions are:

Dirichlet Boundary Condition The boundary constrains a value to the acoustic field itself, i.e. $P(\mathbf{r},\omega)$ is constrained for $\mathbf{r} \in \partial\mathcal{S}$. These boundaries are known as *Dirichlet problems* and their general formulation is:

$$P(\mathbf{r},\omega) = F(\mathbf{r},\omega) \quad \forall \mathbf{r} \in \partial\mathcal{S} \tag{1.15}$$

Usually the field is forced to vanish for $F(\mathbf{r},\omega) = 0\ \forall \mathbf{r} \in \partial\mathcal{S}$, thus it describes a pressure-release boundary.

Neumann Boundary Condition The boundary constrains the normal directional derivative of the acoustic field:

$$\langle \nabla P(\mathbf{r},\omega), \widehat{\mathbf{n}}(\mathbf{r})\rangle = F(\mathbf{r},\omega) \quad \forall \mathbf{r} \in \partial\mathcal{S} \tag{1.16}$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product of two vectors and $\widehat{\mathbf{n}}(\mathbf{r})$ denotes the unit vector normal to $\partial\mathcal{S}$ at a point $\mathbf{r} \in \partial\mathcal{S}$. If the right-hand side of the *inhomogeneous Neumann boundary condition* is forced to vanish, this condition describes a sound hard boundary (e.g. a perfectly reflective wall).

Kirchhoff-Helmholtz Integral Equation Consider the area $\mathcal{S}$ and its boundary $\partial\mathcal{S}$ as depicted in figure 1.1. The boundary value problem is

$$\nabla^2 P(\mathbf{r},\omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r},\omega) = F(\mathbf{r},\omega), \quad \mathbf{r} \in \mathcal{S} \tag{1.17}$$

$$\alpha \langle \nabla P(\mathbf{r},\omega), \widehat{\mathbf{n}}(\mathbf{r})\rangle + \beta P(\mathbf{r},\omega) = 0, \quad \forall \mathbf{r} \in \partial\mathcal{S} \tag{1.18}$$

where $\alpha, \beta \in [0,1]$ account the contributions of the different boundary conditions.

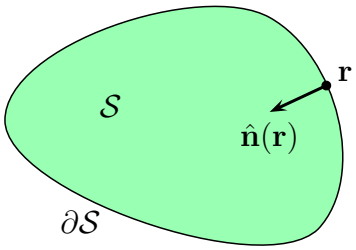


Figure 1.1: Geometry of a boundary value problem in 2D.

The solution is based on the *Huygen's principle*, which states that the wavefront of a propagating wave can be reconstructed by a superposition of spherical waves radiated from every point on the wavefront at a prior instant. Its formulation, the *Kirchhoff-Helmholtz integral*, takes the form

$$a(\mathbf{r})P(\mathbf{r},\omega) = P_0(\mathbf{r},\omega) + \oint_{\partial\mathcal{S}} \left(G(\mathbf{r}|\mathbf{r}',\omega)\left\langle\nabla P(\mathbf{r}',\omega),\hat{\mathbf{n}}(\mathbf{r}')\right\rangle - P(\mathbf{r}',\omega)\left\langle G(\mathbf{r}|\mathbf{r}',\omega),\hat{\mathbf{n}}(\mathbf{r}')\right\rangle\right) d\mathbf{r}' \tag{1.19}$$

where $a(\mathbf{r})$ is a discrimination term:

$$a(\mathbf{r}) = \begin{cases} 1 & \text{for } \mathbf{r} \in \mathcal{S} \\ 0.5 & \text{for } \mathbf{r} \in \partial\mathcal{S} \\ 0 & \text{for } \mathbf{r} \notin \mathcal{S}. \end{cases} \tag{1.20}$$

The equation (1.19) can be interpreted as in figure 1.1. The acoustic field inside the area $\mathcal{S}$ is uniquely determined by three contributions:

1. the acoustic field $P_0(\mathbf{r},\omega)$ due to the source $F(\mathbf{r},\omega)$;
2. the acoustic pressure on $\partial\mathcal{S}$;
3. its directional gradient in the direction normal to $\partial\mathcal{S}$.

The Kirchhoff-Helmholtz integral allows the derivation of physically-motivated acoustic field synthesis methodologies that will be presented in section 1.2.3.

1.2 Sound Field Synthesis Methodologies

This section reviews the key-points of the sound reproduction techniques, starting from the conventional stereophony to the best-known methodologies that receive attention nowadays, *Near-field Compensated Higher Order Ambisonics* (NFC-HOA) proposed in [8], and *Wave Field Synthesis* (WFS) proposed in [9]. The goal of any spatial sound reproduction system is to generate an acoustic field in a way that the listener will perceive the desired sound scene, i.e. he/she will be able to correctly localize all the *virtual* sources and objects, such as walls.

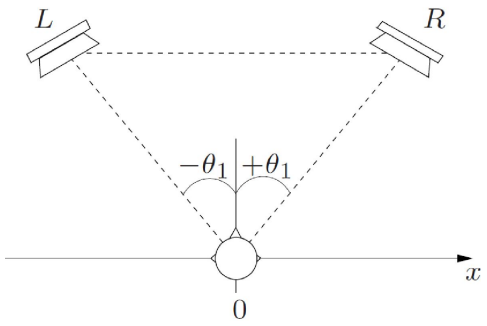


Figure 1.2: The most common configuration for 2-channels stereophony.

1.2.1 Stereophony

The simplest technique is the *two channel stereophony*, which exploits both level and time differences between the two loudspeaker signals. Consider $s(t)$ as the virtual source signal, the loudspeaker signals are

$$\begin{aligned} d_L(t) &= g_L s(t - \tau_L); \\ d_R(t) &= g_R s(t - \tau_R); \end{aligned} \tag{1.21}$$

where g_L, g_R are the gains for the left and right loudspeakers, respectively, and τ_L, τ_R are the time delays for the left and right loudspeakers, respectively.

Assuming the loudspeakers to be in the far field with respect to the listener position, they can be considered as ideal generators of plane waves (described in (1.9)). The most popular loudspeaker arrangement is depicted in figure 1.2; it assumes that the positions of the loudspeakers and that of the listener make up an equiangular triangle. In this setting, the listening position is called *sweet spot* [6]. For any virtual source whose DoA (Direction of Arrival) θ is between $-\theta_1$ and θ_1, it holds that

$$\sin\theta = \sin\theta_1 \frac{1 - g_L/g_R}{1 + g_L/g_R} \tag{1.22}$$

from which one can extract the gain ratio

$$\frac{g_L}{g_R} = \frac{\sin\theta_1 - \sin\theta}{\sin\theta_1 + \sin\theta}. \tag{1.23}$$

This approximation holds only if the loudspeakers and the listener positions are respected, and it exhibits a pronounced sweet spot. Outside the optimality, the spatial perception and also timbral balance degrade. Moreover, the virtual source range is restricted between the two loudspeakers. Indeed, the two loudspeakers act as an angular acoustic window, imposing that $\theta \in [-\theta_1; \theta_1]$.

In order to extend the rendering capabilities to all the directions, many other loudspeakers should be added and placed in order to surround the listener, giving birth to the so called *uniform circular array*. This

is the case of the Vector Based Amplitude Panning (VBAP) technique, which selects the pair of loudspeakers adjacent to the DoA to be rendered and applies to them the gain ratio in equation (1.23).

1.2.2 Ambisonics

The goal of Ambisonics is to extend the panning methods presented above by not using only the loudspeakers closest to the desired source direction. Morover it represents the sound field by a set of signals that are independend on the transducer setup used for recording or reproduction. Ambisonic representation is based on spherical harmonic decomposition of the sound field; in the 2D scenario of interest for the purpose of this work, the pressure field can be written in the cylidrical coordinate system as the *Fourier-Bessel series*, whose terms are the weighted products of directional functions and radial functions

$$P(\rho,\theta) = B_{00}^{+1} J_0(k\rho) + \sum_{m=1}^{\infty} \sqrt{2} J_m(k\rho) \left(B_{mm}^{+1} \cos m\theta + B_{mm}^{-1} \sin m\theta \right), \tag{1.24}$$

where $J_m(\cdot)$ is the Bessel function of order m and B_{mm}^{σ} is the 2D ambisonic component, which can be interpreted as follow: B_{00}^{+1} is the pressure, B_{11}^{+1} and B_{11}^{-1} are related to the gradiented (or also the acoustic velocity) with respect to x and y, respectively.

In practice, let us consider an evenly spaced distribution of N loudspeakers, whose positions are $\mathbf{r}_n = R\left[\cos\theta_n \sin\theta_n\right]^T$ with $n = 0, \ldots, N-1$. The desired sound field, as it woud be generated by the virtual primary sources, is then given as:

$$P(\mathbf{r},\omega) = \sum_{n=0}^{N-1} D_{HOA}(\theta_n, R, \omega) V(\mathbf{r}, \mathbf{r}_n, \omega), \tag{1.25}$$

where $D_{HOA}(\theta_n, R, \omega)$ denotes the driving signal for the n-th secondary source and $V(\mathbf{r}, \mathbf{r}_n, \omega) = \frac{j}{4} H_0^{(2)} \left(\frac{\omega}{c} |\mathbf{r} - \mathbf{r}_n| \right)$ in order to compensate also the near-field instead of using the plane wave assumption of equation (1.11). Finally, the near-field compensated driving functions are:

$$D_{HOA}(\theta_n, R, \omega) = \frac{j}{4} \frac{1}{N} \sum_{\nu=-N'}^{N'} j^{-\nu} \frac{e^{-j\nu\theta_{PW}} e^{j\nu\theta_n}}{H_\nu^{(2)} \left(\frac{\omega}{c} R \right)} \tag{1.26}$$

where ν are the angular frequencies, whose limitation results in a finite series expansion of the plane wave; due to the properties of the Bessel functions, the approximation of the plane wave will be better for low frequencies and smaller distances to origin. For very low orders the wave field will be exact only at the center of the array, limiting the sweet spot. A great variety of optimizations of Ambisonics panning functions has been proposed in literature, e.g. [10], [11].

1.2.3 Wave Field Synthesis

In this subsection we introduce Wave Field Synthesis (WFS) as a physically motivated soundfield reproduction technique. In particular, it is based on the Kirchhoff-Helmholtz integral. Then we illustrate an approach to simplify the implementation of a WFS rendering system adopting only monopole secondary sources easily implemented by loudspeakers on the virtual surface $\partial\mathcal{S}$ as stated by the Huygens principle; figure 1.3 helps visualizing this concept. From a technological point of view, dipole sources (modelled by the directional gradient of the Green function) can be implemented by mounting a loudspeaker on a flat panel, thus they are less efficient. Dipole sources are discarded by applying the Neumann boundary condition to the second addend of the integrand in (1.19)

$$\langle G(\mathbf{r}|\mathbf{r}',\omega), \widehat{\mathbf{n}}(\mathbf{r}')\rangle = \frac{\partial}{\partial\widehat{\mathbf{n}}} G(\mathbf{r}|\mathbf{r}',\omega) = 0 \quad \mathbf{r}' \in \partial\mathcal{S}. \tag{1.27}$$

The solution is known as *Neumann Green function* and it satisfies the boundary condition by taking the form

$$G_N(\mathbf{r}|\mathbf{r}',\omega) = G(\mathbf{r}|\mathbf{r}',\omega) + G(\bar{\mathbf{r}}(\mathbf{r})|\mathbf{r}',\omega), \tag{1.28}$$

where the position $\bar{\mathbf{r}}(\mathbf{r})$ is the mirror image of $\mathbf{r}$ with respect to the tangent plane in $\mathbf{r}'$ on the bound $\partial\mathcal{S}$. Moreover, on the boundary we can write

$$G_N(\mathbf{r}|\mathbf{r}',\omega) = 2G\left(\mathbf{r}|\mathbf{r}',\omega\right), \tag{1.29}$$

which can be inserted as Green function into the Kirchhoff-Helmholz integral (1.19):

$$P(\mathbf{r},\omega) = -\oint_{\partial\mathcal{S}} 2G(\mathbf{r}|\mathbf{r}',\omega)\frac{\partial}{\partial\mathbf{n}} P(\mathbf{r}',\omega)d\mathbf{r}'. \tag{1.30}$$

The result of equation (1.30) is known as the *type-I Rayleigh integral* and it describes the sound field inside an area $\mathcal{S}$ generated by a distribution of monopole sources placed on the boundary $\partial\mathcal{S}$.

Substituting the Green function in (1.13) in the type-I Rayleigh integral (1.30) leads to

$$P(\mathbf{r},\omega) = -\oint_{\partial\mathcal{S}} G_0(\mathbf{r}|\mathbf{r}',\omega)D(\mathbf{r}|\mathbf{r}',\omega)d\mathbf{r}' \quad \mathbf{r} \in \mathcal{S}, \tag{1.31}$$

where $D(\mathbf{r}|\mathbf{r}',\omega)$ denotes the monopole source signals

$$D(\mathbf{r}|\mathbf{r}',\omega) = 2A(\rho)H(\omega)\frac{\partial}{\partial\mathbf{n}} P(\mathbf{r}',\omega). \tag{1.32}$$

Here, $A(\rho) = \sqrt{2\pi\rho}$ and $H(\omega) = \sqrt{\frac{c}{j\omega}}$ are, respectively, a space-dependent attenuation term based on the distance $\rho = \|\mathbf{r} - \mathbf{r}'\|$ and a frequency-dependent term, which are both derived from the far-field Hankel function approximation applied to the 2D Green function:

$$H_0^{(2)}\left(\frac{\omega}{c}\rho\right) \approx \sqrt{\frac{2j}{\pi\left(\frac{\omega}{c}\rho\right)}} e^{-j\left(\frac{\omega}{c}\rho\right)} \tag{1.33}$$

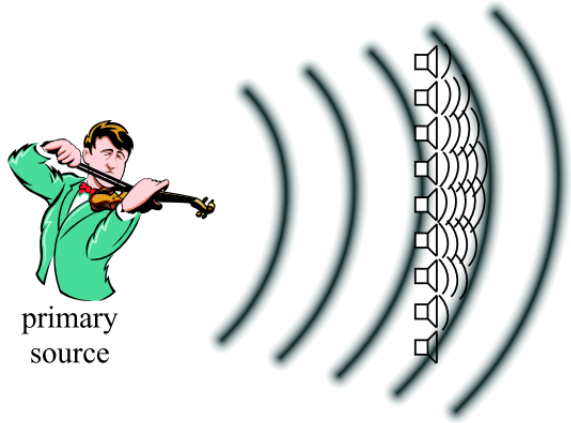


Figure 1.3: The Huygens principle (courtesy of [1]).

$$\tilde{G}(\rho,\omega) = H(\omega)A(\rho)\frac{e^{-j\left(\frac{\omega}{c}\rho\right)}}{4\pi\rho}. \tag{1.34}$$

In conclusion, WFS reproduction is described as a sampling-interpolation process:

- sampling the continuous driving function at the desired secondary source positions, and
- interpolation of the driving function into the listening area.

The spectrum of the secondary sources are not band-limited in the angular frequency domain (i.e. the direction θ) thus the system will not result in an alias-free reproduction. Moreover, no exact reproduction is achieved for nonlinear/planar systems and sampling artifacts exhibit irregular structures.

1.3 Ray Acoustics

In real acoustic propagation problems it may be impractical to solve the wave equation and its boundaries due to the computational costs that would be required. This is the case of realistic room acoustics.

This section introduces the geometrical acoustic starting from the high frequency approximation of the wave equation, which states that the sound wave travels in straight lines that point in the direction of the acoustic energy.

Inherited from the ray optics, this novel paradigm makes use of uniform linear arrays of sensors (microphones) and transducers (loudspeakers) in order to estimate and reproduce the acoustic energy that reaches them from a given set of directions.

1.3.1 The Eikonal Equation

The *ray* is a physical model in which sound is spatially confined and it propagates without angular spread. The sound propagation is thus described as a set of rays traveling in accordance with geometrical rules.

Ray acoustics relies on a set of postulates inherited from ray optics [12]:

1. Light travels in the form of rays emitted by sources and sensed by receivers.

2. In a homogeneous medium (like isotropic air) the speed of propagation c is a function of temperature but can be assumed to be space-invariant. Therefore, the travelling time over a distance d is $\tau = d/c$.

3. Rays do respect the *Fermat's principle* (the travel time is minimized). Therefore, the travel directions are straight lines.

4. The radiance, i.e. the radiant power per unit solid angle per unit projected area, is invariant along the ray.

The validity of these postulates is ensured also for acoustic rays at high temporal frequencies; however this is not verified for real scenarios (e.g. room acoustics) in which the wavelength has the same order of magnitude of the physical dimensions of the acoustic objects (i.e. walls, windows, reflectors, bodies). As a matter of facts, ray acoustics finds applications in simple architectural acoustic problems (e.g. reverberation time estimation, sound focusing systems design).

Ray trajectories can be univocally identified by the surface $\psi(\mathbf{r}, \omega)$, usually referred as *Eikonal*, to which they are normal. Starting from the homogeneous Helmholtz equation (1.6) we can derive its high frequency approximation by rewriting the acoustic field in terms of its magnitude and phase

$$P(\mathbf{r}, \omega) = |P(\mathbf{r}, \omega)| e^{j\psi(\mathbf{r}, \omega)} \tag{1.35}$$

and let $\omega \to \infty$ we obtained the so called *Eikonal equation*

$$\langle \nabla\psi, \nabla\psi \rangle = \left(\frac{\omega}{c}\right)^2. \tag{1.36}$$

The physical interpretation of (1.36) is that it constrains acoustic rays to travel in the direction orhogonal to lines of constant phase.

Let us consider a plane wave and its magnitude-phase factorization:

$$P(\mathbf{r}, \omega) = A(\omega) e^{j\langle \mathbf{k}, \mathbf{r} \rangle} = |A(\omega)| e^{j(\langle \mathbf{k}, \mathbf{r} \rangle + \phi)}. \tag{1.37}$$

In this case the Eikonal becomes

$$\psi(\mathbf{r}, \omega) = \langle \mathbf{k}, \mathbf{r} \rangle + \phi \tag{1.38}$$

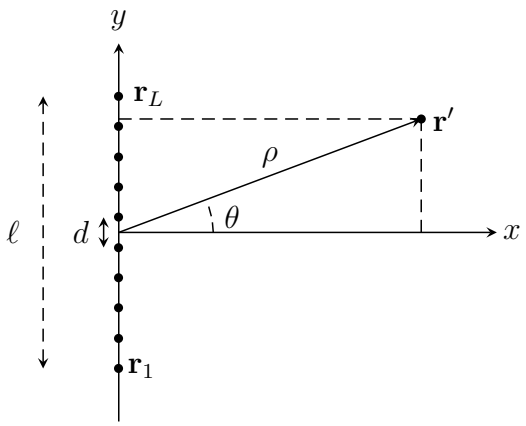


Figure 1.4: Uniform Line Array (ULA).

and its gradient

$$\nabla\psi(\mathbf{r},\omega) = \mathbf{k} \tag{1.39}$$

An immediate interpretation of the results above is that a plane wave is a solution to the Eikonal equation (1.36) as long as the wavenumber vector $\mathbf{k}$ satisfies the dispersion relation (1.10). Moreover, the trajectory of acoustic rays is determined by the gradient of $\psi(\mathbf{r},\omega)$ as described by (1.39).

1.3.2 Spatial Filtering: the Beamformer

The methodologies for analysis and synthesis described in the subsequent chapters rely on spatial filtering. This is a signal processing technique that make use of arrays of sensors or transducers for directional signal detection [13]; originally developed in the electro-magnetic applications such as radar, sonar and radio communications, recently it has been transposed also to acoustics.

The key concept of array processing is the *redundancy* exploited by combining the signals of multiple sensors. In this sense, lots of methods are present in the literature ranging a large set of different applications, such as estimation of the energy distribution [14], source position estimation, source extraction and many others.

These vector signals are processed with the so called *beamformers*, which are spatial filters that applied to the spatial samples of data, enabling the extraction of directional information about the overall acoustic energy.

In the following we consider only arrays of microphones, which are very well studied in the literature; loudspeaker arrays can be considered as the other side of the problem formulation.

Suppose we have L omnidirectional microphones and denote with $s_n(t)$ the signal emitted by the n-th source at time t and with $x_l(t)$ the signal received by the l-the sensor at time t. The latter can be written

as

$$x_l(t) = f_{l,n}(t) * s_n(t) + e_l(t) \tag{1.40}$$

where $f_{l,n}(t)$ denotes a generic transfer function from the source n to the sensor l, $*$ denotes the convolution and $e_l(t)$ is an additive noise, which can be either the electronic noise of the circuits or the background noise of the environment. The Fourier transform of the model above is

$$X_l(\omega) = F_{l,n}(\omega)S_n(\omega) + E_l(\omega). \tag{1.41}$$

By introducing the so called *array transfer vector*

$$\mathbf{f}_n(\omega) = [F_{1,n}(\omega), \ldots, F_{L,n}(\omega)]^T \tag{1.42}$$

we can write the output signal vector as

$$\mathbf{x}(\omega) = \mathbf{f}_n(\omega)S_n(\omega) + \mathbf{e}(\omega) \tag{1.43}$$

where $\mathbf{x}(\omega) = [X_1(\omega), \ldots, X_L(\omega)]^T$ and $\mathbf{e}(\omega) = [E_1(\omega), \ldots, E_L(\omega)]^T$ denotes the additive noise vector. Finally, assuming to have N sources we can extend the previous formula exploiting the superpositon principle, obtaining

$$\mathbf{x}(\omega) = \mathbf{F}(\omega)\mathbf{s}(\omega) + \mathbf{e}(\omega) \tag{1.44}$$

where $\mathbf{F}(\omega) = [\mathbf{f}_1(\omega), \ldots, \mathbf{f}_L(\omega)]^T$ and $\mathbf{s}(\omega) = [S_1(\omega), \ldots, S_N(\omega)]^T$.

1.3.2.1 Far Field sources

The expression of the transfer function from the n-th source to the l-th microphone depends on the distance of the sources from the array. If sources are sufficiently far, the wavefronts can be reasonably modelled as plane waves. For linear arrays, to be considered in far field the sources should respect the following

$$\rho > \frac{2\ell}{\lambda} \tag{1.45}$$

where ℓ denotes the length of the array.

In this context, the transfer function assumes the form of

$$F_{l,n}(\omega) = e^{-j\langle \mathbf{k}_n, \mathbf{r}_l \rangle} \tag{1.46}$$

where, limited to a 2D scenario, $\mathbf{k}_n = \frac{\omega}{c}[\cos\theta_n, \sin\theta_n]^T$ is the wavenumber vector of the n-th source and $\mathbf{r}_l = [x_l, y_l]^T$ is the position of the l-th microphone, thus θ_n denotes the DoA of the n-th source.

Although many other geometries are present in the literature, this work of thesis considers only the uniform linear array, i.e. L microphones uniformly spaced at distance d along the y axis and centered in the origin as depicted in figure 1.4.

Under these assumptions, the acoustic transfer function becomes:

$$F_{l,n}(\omega) = e^{-j\frac{\omega}{c}y_l \sin\theta_n} \tag{1.47}$$

$$\mathbf{f}_n(\omega) = \left[e^{-j\frac{\omega}{c}y_1 \sin\theta_n}, \ldots, e^{-j\frac{\omega}{c}y_L \sin\theta_n}\right]^T \tag{1.48}$$

and, exploting the so called *central sensor as reference* without loss of generality

$$\mathbf{f}_n(\omega) = \left[e^{j\frac{\omega}{c}\sin(\theta_n)d\left(\frac{L-1}{2}\right)}, \ldots, 1, \ldots, e^{-j\frac{\omega}{c}\sin(\theta_n)d\left(\frac{L-1}{2}\right)}\right]^T \tag{1.49}$$

Spatial aliasing condition Similarly to the temporal Shannon sampling theorem, a spatial aliasing condition is derived by defining a *spatial frequency*

$$\omega_s = \omega\frac{d\sin\theta}{c} \tag{1.50}$$

and rewriting the array transfer vector (1.49) as

$$\mathbf{f}_n(\omega) = \left[e^{j\omega_s\left(\frac{L-1}{2}\right)}, \ldots, 1, \ldots, e^{-j\omega_s\left(\frac{L-1}{2}\right)}\right]^T. \tag{1.51}$$

We have to constrain the spatial frequency in order to avoid spatial aliasing:

$$|\omega_s| \leq \pi \rightarrow |\omega\frac{d\sin\theta}{c}| \leq \pi \rightarrow \omega\frac{d|\sin\theta|}{c} \leq \pi \tag{1.52}$$

that in the worst case becomes

$$\omega\frac{d}{c} \leq \pi \rightarrow 2\pi\frac{d}{\lambda} \leq \pi \tag{1.53}$$

yielding the final condition:

$$d \leq \frac{\lambda}{2} \quad \forall\theta \tag{1.54}$$

Finally, if the condition (1.45) does not hold and the plane waves assumption is no longer valid, the transfer function $F_{l,n}(\omega)$ can be written similarly to the Green function in (1.13):

$$F_{l,n}(\omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}_l - \mathbf{r'}_n\|}}{4\pi\|\mathbf{r}_l - \mathbf{r'}_n\|} \tag{1.55}$$

1.3.2.2 Delay-and-sum beamformer (DAS)

In this section we present a simple but powerful beamformer design procedure known in the literature as *delay-and-sum* (DAS).

Let us consider a uniform linear microphone array deployed along the y axis and centered in the origin and a source in far-field placed in $\mathbf{r'} = \rho_s[\cos\theta_s, \sin\theta_s]^T$ that emits a signal $s(t)$. Assuming the model of equation (1.43), our purpose is to extract the signal by applying a FIR filter to the output of the array signal:

$$\hat{s}(\omega) = \mathbf{h}^H(\omega)\mathbf{x}(\omega), \tag{1.56}$$

where $\mathbf{h}^H(\omega) = [h_1(\omega), \ldots, h_L(\omega)]^H$ is the vector of filter coefficients. Let us define the variance of the output as:

$$\mathbf{E}\left\{|y(\omega)|^2\right\} = \mathbf{h}^H(\omega)\mathbf{\Phi}_{\mathbf{xx}}(\omega)\mathbf{h}(\omega) \tag{1.57}$$

where $\mathbf{\Phi}_{\mathbf{xx}}(\omega) = \mathbf{E}\left\{\mathbf{x}(\omega)\mathbf{x}^H(\omega)\right\}$ is the auto-covariance matrix of the array signal.

Therefore, we can set up a minimization problem as follows:

$$\begin{aligned} \underset{\mathbf{h}(\omega)}{\arg\min} \quad & \mathbf{h}^H(\omega)\mathbf{A}(\omega)\mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{h}^H(\omega)\mathbf{B}(\omega) = \mathbf{c}(\omega) \end{aligned}$$

where $\mathbf{A}(\omega) \in \mathbb{C}^{L\times L}$,$\mathbf{B}(\omega) \in \mathbb{C}^{L\times q}$ and $\mathbf{c}(\omega) \in \mathbb{C}^{1\times q}$ are generic frequency-dependent complex-valued matrices and q is the number of constraints. Omitting the frequency dependence, the generic solution of this optimization problem is given by

$$\mathbf{h}_o = \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{B}^H\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{c}^H. \tag{1.58}$$

In the specific case of the DAS beamformer, assuming the signal $\mathbf{x}(\omega)$ to be spatially white implies that $\mathbf{A}(\omega) = \mathbf{\Phi}_{\mathbf{xx}}(\omega) = \mathbf{I}$. Moreover, constraining the beamformer to pass undistorted the source signal is exploited by using one single constraint defined as

$$\mathbf{h}^H(\omega)\mathbf{f}(\omega) = 1, \tag{1.59}$$

where $\mathbf{f}(\omega)$ collects the acoustic transfer function from the source to each array element.

Finally, the DAS optimization problem takes the form

$$\begin{aligned} \underset{\mathbf{h}(\omega)}{\arg\min} \quad & \mathbf{h}^H(\omega)\mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{h}^H(\omega)\mathbf{f}(\omega) = 1, \end{aligned}$$

yielding to the solution

$$\mathbf{h}_o(\omega) = \frac{\mathbf{f}(\omega)}{\|\mathbf{f}(\omega)\|^2}. \tag{1.60}$$

As a simple interpretation, the DAS beamformer applies a delay to each microphone making sure that the desired signal is summed constructively. Finally, the DAS beamformer is completely data-independent, thus it requires a small effort from a computational stand point.

1.4 Plenacoustic Imaging

In this section we introduce the reader to the realm of Plenacoustic imaging, or *soundfield imaging*, as a method for acoustic scene analysis originally presented in [3]. Its main advantages are that the images are generated by a common processing layer and can be processed using methods inherited from pattern analysis literature.

The name is inherited from the concept of *plenacoustic function* (PAF) , that describes the acoustic radiance in every direction through every point in space [15]. In a 2D geometric domain, it can be written as a five-dimensional function $f(x, y, \theta, \omega, t)$ of position (x, y), direction θ, frequency ω and time t.

In [2], the soundfield $p(\mathbf{r}, \omega)$ is expressed as the superposition of plane waves with wave number vector $\mathbf{k}(\theta)$:

$$p(\mathbf{r}, \omega) = \frac{1}{2\pi} \int_0^{2\pi} e^{j\langle \mathbf{k}(\theta), \mathbf{r} \rangle} \tilde{P}\left(\mathbf{k}(\theta)\right) d\theta, \tag{1.61}$$

where $\tilde{P}\left(\mathbf{k}(\theta)\right)$ is known as *Herglotz density function* and it modulates in amplitude and phase each plane wave component. The plenacoustic function is defined as the integrand of (1.61)

$$f(x, y, \theta, \omega) \triangleq e^{j\langle \mathbf{k}(\theta), \mathbf{r} \rangle} \tilde{P}\left(\mathbf{k}(\theta)\right), \tag{1.62}$$

where the time dependencies have been omitted since we are particularly interested in the dependence on the position, direction and frequency.

In [3] the authors captures the PAF by means of microphone arrays that act as Observation Windows (OWs) for the acoustic scene. This approach consists of dividing the microphone array into subarrays, and applying the the plane-wave analysis on individual subarrays. Each directional component is obtained through beamforming techniques that allows to scan the acoustic field for a discrete set of directions. In other words, the emerging image represents an estimate of the acoustic energy carried by directional components measured at several points on the array.

The subdivision into sub-arrays implies an important consideration: for the far-field assumption to be valid, sources no longer need to be in the far field with respect to the length of the global array, but only with respect to the size of the sub-arrays. On the other hand, performances degrate at low-frequencies and when the distance becomes comparable to the length of the sub-arrays.

1.4.1 The Reduced Euclidean Ray Space

In [2] a novel parametric parametrization of the *ray space* has been introduced as the domain of the plenacoustic function, in which each value associated to a ray is represented as a point in this space.

Under the radiance invariance law assumption, we can establish an equivalence between rays and oriented lines. In the 2D scenario:

$$l_1 x + l_2 y + l_3 = 0, \tag{1.63}$$

or, in vector notation,

$$\mathbf{p}^T \mathbf{l} = 0, \quad \mathbf{l} = [l_1, l_2, l_3]^T, \quad \mathbf{p} = [x, y, 1]^T. \tag{1.64}$$

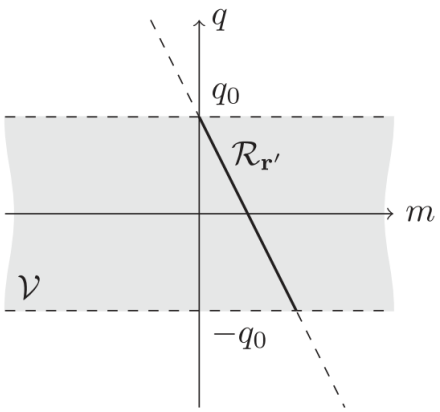


Figure 1.5: Ray Space representation of a point source in $\mathbf{r}'$.

From the latter we can state that all the vectors of the form $\mathbf{l} = k[l_1, l_2, l_3]^T$, $k \neq 0$ represent the same line and form a class of equivalence. Thus, the vector $\mathbf{l}$ can represent all rays with arbitrary direction.

Hence, we can think of a domain with coordinates l_1, l_2, l_3 where the vector $\mathbf{l}$, corresponding to a ray, is a point. This domain will be referred as *Projective Ray Space* $\mathcal{P}$. Moreover, in order to distinguish between two oppositely oriented rays lying on the same line, we can constrain k to either positive or negative values only, defining the *oriented* projective space.

A particular reduced ray space, called the *Euclidean ray space* (m, q), is derived from the projective ray space by setting

$$\begin{aligned} m &= -\frac{l_1}{l_2} \\ q &= -\frac{l_3}{l_2}. \end{aligned} \tag{1.65}$$

In [3] the authors provide a detailed study of how the acoustic primitives can be represented in this reduced space. As we can see from equation (1.65), rays parallel to the y axis can not be represented in a limited domain; deploying only one observation window along the y axis overcomes this limitation. Moreover, since we loose the orientation of the rays, we conventionally assume the orientation toward the y axis from the positive half-plane $x > 0$.

Figure 1.5 shows how a point source looks like in this reduced ray space for an OW deployed along the y axis from $-q_0$ to $+q_0$. The image of a point source becomes a line in (m, q) whose parameters are the coordinates of the point. The visibility $\mathcal{V}$ of the array appears as a stripe between the limits$-q_0$ and $+q_0$ of the OW:

$$\mathcal{V} = \{(m, q) \in \mathbb{R} \times \mathbb{R} : -q_0 \leq q \leq q_0\}. \tag{1.66}$$

The Euclidean parametrization introduced in [3] has been used in [4] to devise an important tool called *Ray Space Transform* that performs a

transformation of the signals captured by a microphone array based on a short space-time Fourier transform using discrete Gabor frames. The Ray Space Transform will be analysed in detail in the following chapter.

1.5 Conclusive Remarks

In this chapter we introduce the main issues in modelling the sound field and propagation phenomena.

First we have addressed the so called nonparametric representation, in which the acoustic field is described as a function of space and time and it is decomposed using different basis functions, foremost the planes wave decomposition. Then we have reviewed the principal methodologies for soundfield reproduction, starting from the simplest two-channel stereophony up to the more immersive techniques such as Ambisonics, which is based on a spherical harmonics decomposition of the wave field, and Wave Field Synthesis, which can be described as as a sampling-interpolation process of the continuous driving function that emerges from the Kirchhoff-Helmholtz integral. These techniques are strongly motivated in multimedia applications such as gaming or telepresence; indeed, the immersivity of a sound experience relies on the capability of rendering realistic sources and propagations. However, in addition to their peculiar limitations, these nonparametric methods result in a limited sweet spot, requiring the listener to be in a specific position, e.g. inside a circular array for Ambisonics.

In the second part we have presented the plenacoustic techniques that are emerging in the last decade. The acoustic field is modelled as the superpositions of acoustic rays, propagating without angular spread. By means of uniform linear arrays of sensors and transducers, the core processing of these techniques is the beamformer, a spatial filter that scans the sound field in a given set of directions. The output is a image that can be processed with algorithms inherited from pattern analysis literature.

2
Theoretical Background

This chapter introduces the readers to the linear operator known as *Ray Space Transform*, which is the most important tool this work of thesis relies on for the analysis of sound scenes, and the concept of *radiation pattern* as the angular-frenquency dependence of any acoustic object.

We start describing how a sound scene can be acquired by means of microphone arrays and analysed in a plenacoustic fashion. In plenacoustic imaging, acoustic primitives such as sources, observation windows and reflectors appear as linear patterns in the resulting images, thus enabling the use of pattern analysis techniques to solve acoustic problems. Briefly, the plenacoustic function (1.62) is estimated by subdividing the microphone array into smaller overlapped sub-arrays and by estimating each directional component for each sub-array through beamforming. The main limitation of this methodology is that performance degradates at lower frequencies and when the distance of the source becomes comparable to the size of the sub-array. In order to overcome this limitation, the authors of [4] propose a new invertible transformation in order to map the extracted directional information onto the reduced Euclidean ray space described in 1.4.1. This new linear operator, which takes the name of *Ray Space Transform*, shows perfect reconstruction capabilities and therefore is a good candidate for sound scene processing paradigms.

The second section describes in detail the radiation pattern as an amplitude function of both the temporal frequency and the angles with respect to a reference direction. Inherited from the literature on antennas and electro-magnetic devices, the problem of modelling the radiation pattern has been investigated by many authors. In the audio equipment industry, the polar data are acquired by rotating the loudspeaker or the microphone on a turntable and measuring its output at each turntable

position [16]. The most common representation, because of its physical interpretation, is the Spherical Harmonics Decomposition, and its 2D formulation known as *Circular Harmonics Decomposition* (CHD). In both the electro-magnetic and acoustic fields, the radiation pattern is strongly related to the physical characteristics of the source, such as the dimensions and the materials. Thus it includes information not only about the position and orientation of the source, but, in a way, it can give a flavour of the characteristic of a object in the scene.

2.1 Ray Space Transform

In [4] the sound field decomposition relies on a new overcomplete basis of wave functions of local validity. In other words, a local space-time Fourier analysis is performed by computing the similarity between the array data and shifted and modulated copies of a prototype spatial window.

In the time domain, the local Fourier transform of a signal $s(t)$ is defined as the discrete Gabor expansion [17]

$$[\mathbf{G}]_{i,w} = \int_{\mathbb{R}} s(t)\psi^*_{i,w}(t)dt. \tag{2.1}$$

where

$$\psi^*_{i,w}(t) = \psi(t - iT)e^{-j\frac{2\pi}{W}w(t-iT)}. \tag{2.2}$$

$\psi(t)$ is the analysis window, $i \in \mathbb{Z}$ is the time frame index, $w = 0, \ldots, W-1$ is the frequency band index, W is the number of frequency bins and T is the window hop size.

From $\mathbf{G}$ is possible to reconstruct the time signal $s(t)$ using the local inverse Fourier transform defined as

$$s(t) = \sum_i \sum_{w=0}^{W-1} [\mathbf{G}]_{i,w}\, \tilde{\psi}_{i,w}(t), \tag{2.3}$$

where $\tilde{\psi}_{i,w}(t) = \tilde{\psi}(t - iT)e^{-j\frac{2\pi}{W}w(t-iT)}$ and $\tilde{\psi}(t)$ being the synthesis window; to achieve perfect reconstruction, the completeness condition

$$\sum_i \psi(t - iT)\tilde{\psi}(t - iT) = \frac{1}{W} \tag{2.4}$$

must be satisfied. The analysis and synthesis windows ψ and $\tilde{\psi}$ form a pair of dual discrete Gabor frames [18].

The above Gabor frame definitions can be extended to the ray space domain. First, let us consider a continuous observation window along the y axis limited from $-q_0$ and q_0, as done in section 1.4.1, and exploit the plane-wave expansion to model the sound field. We recall the reader that, given the wavenumber $\mathbf{k} = [\cos\theta, \sin\theta]^T$, the pressure field is:

$$p(\mathbf{r}, \omega) = e^{-j\frac{\omega}{c}\langle \mathbf{k}(\theta), \mathbf{r}\rangle}. \tag{2.5}$$

The expression above can be reduced to the observation window along the y axis:

$$p(\mathbf{r}, \omega) = e^{-j\frac{\omega}{c} y \sin\theta}. \tag{2.6}$$

.

Starting from the classical parametrization of the plenacoustic function $f(x, y, \theta)$ (1.62), a new mapping is defined by

$$\begin{cases} x = 0 \\ \theta = \arctan(m) \quad -\pi/2 < \theta < \pi/2 \\ q = y. \end{cases} \tag{2.7}$$

Note that the first condition is derived by placing the OW along the y axis. The phase shift at position y due to a directional contribution from θ is given by

$$y \sin\theta = y \sin(\arctan m) = \frac{ym}{\sqrt{m^2+1}}, \tag{2.8}$$

where the second equality follows from known interrelations among trigonometric functions [19].

We adopt an uniform grid for sampling the (m, q) plane and denote by $\bar{q}$ and $\bar{m}$ the sampling intervals on the q and m axes, respectively. In particular, we choose

$$\begin{aligned} q_i &= \bar{q}\left(i\frac{I-1}{2}\right), \qquad i = 0, \ldots, I-1 \\ m_w &= \bar{m}\left(w\frac{W-1}{2}\right), \quad w = 0, \ldots, W-1 \end{aligned} \tag{2.9}$$

I and W being the number of samples on q and m axes, respectively, and we adopt the Gaussian window

$$\psi(y) = e^{-\pi y^2/\sigma^2} \tag{2.10}$$

where the scalar σ controls the width of the window. Thus, we can write the Ray Space Transform (RST) of the acoustic field $P(y, \omega)$ in the continuous setting as follows

$$[\mathbf{Y}]_{i,w}(\omega) = \int_{-q_0}^{q_0} P(y, \omega) e^{-j\frac{\omega}{c}\frac{ym_w}{\sqrt{m_w^2+1}}} \psi_{i,w}^*(y) dy. \tag{2.11}$$

The Inverse Ray Space Transform (IRST) is defined in the same way as (2.3) in the time domain, yielding

$$P^{(\mathbf{Y})}(y, \omega) = \sum_{i=0}^{I-1} \sum_{w=0}^{W-1} [\mathbf{Y}]_{i,w}(\omega) e^{j\frac{\omega}{c}\frac{ym_w}{\sqrt{m_w^2+1}}} \tilde{\psi}_{i,w}^*(y), \tag{2.12}$$

where $\tilde{\psi}_{i,w}(y)$ is the synthesis window.

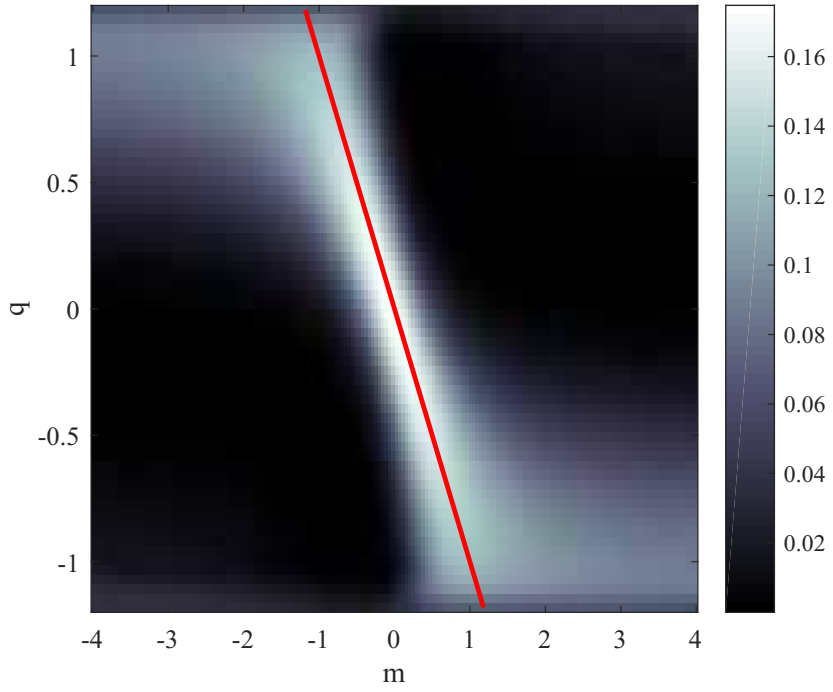


Figure 2.1: (m, q) line (in red) and magnitude of the RST for an isotropic point source, emitting a pure tone at 2 kHz, placed in $\mathbf{r}' = [1, 0]$ m.

In real scenario we do not have a continuous aperture but we sample the spatial signal in a finite number of points in the space. For this purpose, let us consider a uniform linear array of L microphones displaced along the y axis between $-q_0$ and q_0 and centered in the origin, thus the l-th microphone is placed in

$$y_l = d\left(l - \frac{L-1}{2}\right), \quad l = 0, \ldots, L-1 \tag{2.13}$$

where d is the distance between adjacent microphones. After discretizing (2.11) with respect to y we obtain

$$[\mathbf{Y}]_{i,w}(\omega) = d\sum_{l=0}^{L-1} P(y_l, \omega) e^{-jk\frac{y_l m_w}{\sqrt{m_w^2+1}}} e^{-\pi\frac{(y_l-q_i)^2}{\sigma^2}}. \tag{2.14}$$

The *discrete ray space transform* can be conveniently written in matrix form by introducing the matrix $\mathbf{\Psi}(\omega) \in \mathbb{C}^{IW\times L}$ whose element in row l and column $(i + wI + 1)$ is

$$[\mathbf{\Psi}]_{l,i+wI+1} = e^{-jk\frac{y_l m_w}{\sqrt{m_w^2+1}}} e^{-\pi\frac{(y_l-q_i)^2}{\sigma^2}} d. \tag{2.15}$$

We also introduce the *canonical dual matrix* $\tilde{\mathbf{\Psi}} \in \mathbb{C}^{IW\times L}$ corresponding to the pseudo-inverse of $\mathbf{\Psi}$

$$\tilde{\mathbf{\Psi}} = \left(\mathbf{\Psi}\mathbf{\Psi}^H\right)^{-1}\mathbf{\Psi}, \tag{2.16}$$

which, among all the infinite matrices that could play the role of dual matrices of $\mathbf{\Psi}$ (due to the overcompleteness condition $IW \geq L$), is the one that guarantees the coefficients to have minimum norm.
Finally we introduce also the vector $\mathbf{y} \in \mathbb{C}^{IW \times 1}$ obtained by rearranging the elements of $\mathbf{Y}$ as

$$[\mathbf{y}]_{i+wI+1} = [\mathbf{Y}]_{i,w}\,. \tag{2.17}$$

Therefore, we can write the discrete ray space transform as

$$\mathbf{y} = \mathbf{\Psi}^H \mathbf{p} \tag{2.18}$$

and its inverse

$$\mathbf{p}^{(\mathbf{Y})} = \tilde{\mathbf{\Psi}}^H \mathbf{y}. \tag{2.19}$$

In order to give the reader a graphical interpretation, figure 2.1 shows the linear pattern and the magnitude of the discrete ray space coefficients of the acoustic field of a point source in $\mathbf{r}' = [1, 0]$ meters, emitting a tone at 1kHz. For the sake of completeness, the parameters of the RST are $\sigma = 22.5$cm, $L = 32$ microphones with $d = 7.5$cm ($q_0 = 1.1625$m), $\bar{q} = d/2 = 3.75$cm and $\bar{m} = 4.47$cm.

2.1.1 Remarks

Notice that the array has, inevitably, a limited extension, bounded within $-q_0$ and q_0. Therefore, in addition the the moving spatial window (2.10), there is a fixed rectangular window of length $2q_0$.

The interpretation of (2.11) is immediate upon considering a single spatial window, i.e. upon fixing i. The expression is intepreted as the beamforming operation applied to the aperture data that have previously been weighed by a Gaussian spatial windowing function centered at q_i. Thus, $[\mathbf{Y}]_{i,:}$ collects the outputs of multiple beamforming operations, each computed from a specifically weighed portion of the aperture data.

2.1.2 Source Localization in the Ray Space

In this section we discuss one example of application, where the analysis and processing of the acoustic signal are done in the ray space presented above. The example is taken from [4], where the author aims at localizing acoustic sources in near field condition.

Localization is based on a wide band extension of the ray space coefficients. Denoting with $|\mathbf{Y}(\omega_k)|$ the magnitude of the ray space coefficients at frequency ω_k, the wideband magnitude estimated is defined as

$$\mathcal{Y} = \left(\prod_{k=0}^{K/2} E\left\{ |\mathbf{Y}(\omega_k)| \right\} \right)^{2/K}, \quad \mathcal{Y} \in \mathbb{R}^{I \times W}, \tag{2.20}$$

where $K/2$ is the number of considered frequencies. In the simulations the expectation is approximated by an average over $J = 10$ time frames:

$$E\{\mathbf{Y}(\omega_k)\} = \frac{1}{J}\sum_{j=0}^{J-1}\mathbf{Y}^{(j)}(\omega_k), \tag{2.21}$$

where $\mathbf{Y}^{(j)}(\omega_k) = \mathbf{\Psi}^H(\omega_k)\mathbf{u}^{(j)}(\omega_k)$ are the ray space coefficients computedt at the j-th time frame from the array signal $\mathbf{u}$. As stated in section 1.4.1 the image of a point source in the reduced ray space (m, q) is a line whose parameters are the (x, y) coordinates of the point source. Therefore, localization is accomplished by estimating the peaks for each row $\mathcal{Y}_{i,:}$. The position of the source is estimated through a least-squares regression of the locations of the peaks. In particular, the founded peaks at value $\hat{m}_i$ are collected in the vector $\hat{\mathbf{m}} = [\hat{m}_0, \ldots, \hat{m}_{I-1}]^T$; then we introduce a matrix $\hat{\mathbf{M}} = [-\hat{\mathbf{m}}, \mathbf{1}]$ and the vector $\mathbf{q} = [q_0, \ldots, q_{I-1}]^T$ that collects the position of the centres of the spatial windows on the y axis. Finally, since $\hat{\mathbf{M}}\mathbf{r}' = \mathbf{q}$ the position $\mathbf{r}'$ is estimated as

$$\hat{\mathbf{r}'} = \left(\hat{\mathbf{M}}^T\hat{\mathbf{M}}\right)\hat{\mathbf{M}}^T\mathbf{q}. \tag{2.22}$$

2.2 Radiation Pattern

This section introduces the concept of *radiation pattern* as a angular-frequency dependence of the amplitude of a signal emitted/received by a source/microphone. In general, acoustic sources present emission patterns which are far from being omnidirectional, i.e. constant with respect to the directions, in particular at higher frequencies [20]. Considering a loudspeaker, for example, the complexity of its radiation pattern at a given frequency depends on the ratio between the size of the source and the considered wavelength [21]. As a matter of facts, the radiation pattern influences the way the acoustic waves propagate within a sound scene [22], [20], with direct implications on the behavior of the most audio processing algorithms.

Many authors in the literature assume the radiation pattern to be constant over both the frequency and the directions of propagation, thus either simplifying the algorithms and the results or introducing an equalization stage, e.g. [23]. Later, some authors have assumed arbitrary radiation directivities in their soundfield rendering algorithms (e.g. in Ambisonics [24] and in Wave Field Synthesis [25], [26]).

The rest of this chapter is organized as follows. First we give some formal definitions of radiation pattern, along with the necessary assumptions for the correcteness with respect to the acoustic propagation laws. Then we present a parametric approach to model the radiation pattern in a efficient way, inherited from the literature on antennas and EM propagation. This method decomposes the radiation function as a sum of cosines of multiples of the directions, the so-called *circular harmonics*.

Finally, we give an example of the advantages that soundfield imaging techniques can bring for the purpose of analysing the radiation pattern of complex musical instruments.

2.2.1 Definition

The directivity pattern describes the response as a function of the direction of propagation in a specified plane and at a specified frequency. It is defined starting from the far field solution of the Rayleigh's first integral [7].

Consider a primary source in the xy-plane; the pressure will be given by

$$P(\mathbf{r}, \omega) = S(\omega)D(\theta, \omega)g(\mathbf{r}, \omega), \tag{2.23}$$

where $S(\omega)$ is the temporal Fourier transform of the source signal and $D(\theta, \omega)$ is the directivity function of the source in the far field approximation and $g(\mathbf{r}, \omega)$ is the Green function:

$$g(\mathbf{r}, \omega) = \frac{e^{-jk\rho}}{4\pi\rho} \tag{2.24}$$

being ρ the distance from the source such that $k\rho \gg 1$, k being the wavenumber. The far-field directivity pattern is defined by removing the Green function $g(\mathbf{r}, \omega)$ from the equation (2.23), so that directivity pattern $D(\theta, \omega)$ is defined by

$$\lim_{\rho\to\infty} P(\mathbf{r}, \omega) = \frac{e^{-jk\rho}}{\rho} D(\theta, \omega). \tag{2.25}$$

From equation (2.23) we can state that the radiation pattern is the absolute value of the directivity function, i.e. $|D(\theta, \omega)|$ and describes the intensity of the sound field emitted towards direction θ at frequency ω. For omnidirectional (isotropic) point source $|D(\theta, \omega)| = 1$.

2.2.2 Circular Harmonics Decomposition

This section introduces the parametrization of the radiation pattern through spherical harmonics, as done in [21]. Since the scope of this work is the analysis and the synthesis of acoustic fields restricted to a plane, the spherical harmonics model is reduced into its 2D formulation, called *Circular Harmonics Decomposition*. In [27] the authors show that the sound field in a whole listening area can be reconstructed starting from a relatively small number of sampling points on a circumference.

In order to obtain such a modal representation of the sound field, we need to adopt a polar coordinate system (radius ρ and angle θ). Thanks to the 2π-periodicity of $P(\rho, \theta, \omega)$ with respect to the angle θ, it can be written as a Fourier series with angular coefficients $\dot{P}_\mu(\rho, \omega)$:

$$\mathcal{S}_\theta \{P(\rho, \theta, \omega)\} = \dot{P}_\mu(\rho, \omega) = \frac{1}{2\pi} \int_0^{2\pi} P(\rho, \theta, \omega) e^{-j\mu\theta} d\theta. \tag{2.26}$$

If we consider a spatial region free of sources, we can write the angular coefficients as

$$\dot{P}_\mu(\rho, \omega) = C_\mu(\omega) J_\mu(k\rho), \tag{2.27}$$

where $C_\mu(\omega)$ is the μ-th circular harmonic at frequency ω and $J_\mu(k\rho)$ are the Bessel functions of the first kind and order μ. This kind of modelling efficiently encodes complex radiation patterns: each frequency can be represented completely by the coefficients, independently on the angular domain. Moreover, this allows to fastly reconstruct the value of the radiation pattern at arbitrary directions.

2.2.3 Radiation Pattern and Soundfield Imaging

In [28], the authors propose a methodology for measuring the radiation pattern of a violin played by a violinist. In the literature, several methodologies have been proposed for characterizing the spatial cues of this complex instrument, either based on artificial excitation mechanisms (i.e. without any violinist) or in anechoic rooms where the player rotates in front of a microphone array. However, most of these techniques prevent the musician to play in a natural fashion, due to both the absence of reflections in the room and the fixed positions he/she must keep.

The proposed methodologies relies on the plenacoustic analysis, determining directional components of the sound field at multiple point in space. Being $P(\mathbf{x}, \mathbf{u}, \omega) \in \mathbb{C}$ the plane wave component propagating towards direction $\mathbf{u}$ and contributing to the sound field in $\mathbf{x}$, it holds that:

$$P(\mathbf{x}_i, \mathbf{u}_i, \omega) = g(\mathbf{r}_i, \omega) S(\omega) D(\theta_i, \omega), \tag{2.28}$$

where the point $\mathbf{x}_i$ is a point in the space belonging to an evaluation grid, $\mathbf{u}_i$ is the direction of the ray originating at $\mathbf{x}_s$ (the violin) and passing through $\mathbf{x}_i$. Thus the radiation pattern is easily obtained as

$$|D(\theta_i, \omega)| = \|\mathbf{x}_i - \mathbf{x}_s\| \frac{|P(\mathbf{x}_i, \mathbf{u}_i, \omega)|}{|S(\omega)|}. \tag{2.29}$$

2.2.4 Remarks

In the second half of this chapter we have presented the radiation pattern as a directivity function, i.e. the amplitude of the sound field as a function of both the travelling directions and the frequency (due to the relation between the wavelength and the physical dimensions of the acoustic object). This descriptor has a crucial importance for the purpose of analysing and manipulating acoustic scenes, both for characterizing any sound object and for retrieving its orientation in space. We have introduced a parametric model, the Circular Harmonics Decomposition, that represents the sound field as a Fourier series, thanks to the periodicity in the spherical coordinate system. This way, the radiation pattern

can be efficiently encoded and thus reconstructed with a small number of coefficients, independently on the angular domain.

Finally we have shown an interesting application of the plenacoustic imaging for the extraction of the radiation pattern of a violin during the performance. The framework presented in [28] is the solid root of this work of thesis.

3

Acoustic Scene Analysis and Manipulation

In this chapter we present the sound scene manipulation methodology devised in this thesis. This work relies on the parametric spatial sound processing techniques that are emerging in the literature [29], [30]. The main idea of this framework is to represent an input audio scene in a way that is independent of any assumed or intended reproduction format, thus enabling optimal reproduction over any given playback system as well as flexible scene modification [31].

We have adopted the Ray Space Transform framework of [4] in order to extract all the information about the scene geometry. In particular, the acoustic scene is acquired by a uniform linear array of microphones, whose signal is transformed in the reduced ray space. The source localization is efficiently performed thanks to the linear pattern that emerges from the plenacoustic image, as shown in section 1.4.1.

The radiation pattern is estimated starting from the ray space image compensated for the distance between the array and the source. This estimate is encoded with the Circular Harmonics Decomposition presented in 3.1.2.3 in order to estimate the small set of coefficients that enable an extremely efficient parametrization of the source orientation and directivity. Then a specifically-designed beamformer extracts the source signal.

Finally we present the sound field reproduction based on a uniform linear array. Although in this work the synthesis stage is very similar to the analysis stage, the acoustic descriptors provided to the reproduction system is format-agnostic.

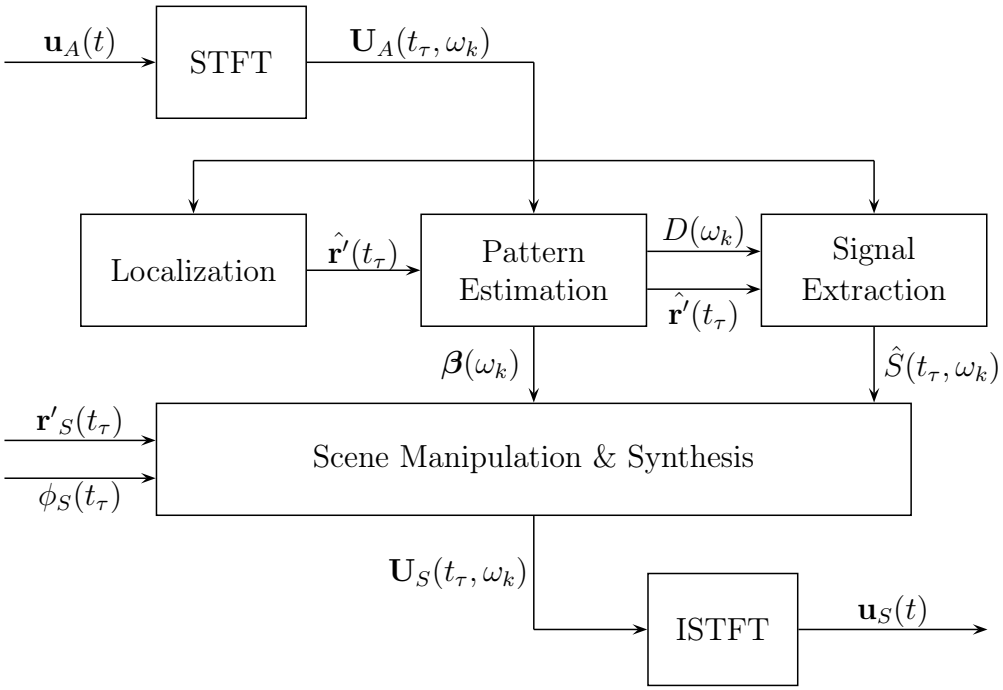


Figure 3.1: Block diagram for the whole manipulation system.

3.1 Scene Analysis

This section describes the spatial sound scene acquisition, adopting the parametric representation. First we present the acoustic descriptors that allows to encode a sound scene in a compact and efficient way. Then, we propose a new methodology for extracting and encoding the radiation pattern and the source signal. Figure 3.1 shows the main steps at a high level perspective.

The sound scene analysis is performed in a plenacoustic fashion starting from a uniform linear array of L_A omnidirectional microphones, whose elements distance is d_A; the microphone positions are:

$$[x_l, y_l]^T = \left[0, \quad d_A \left(l - 1 - \frac{L_A}{2}\right)\right]^T \quad l = 1, \ldots, L_A. \tag{3.1}$$

Scene Geometry As depicted in figure 3.2 the positions and the dimensions of the acoustic objects in the scene are described in a Cartesian coordinate system, which originates in the center of the microphone array. In particular, the y axis is oriented along the array and the x axis is orthogonal to the array. However, the most convenient representation of the source position relies on the polar coordinate system; indeed, the acoustic propagation is conveniently expressed in terms of directions θ and distances ρ.

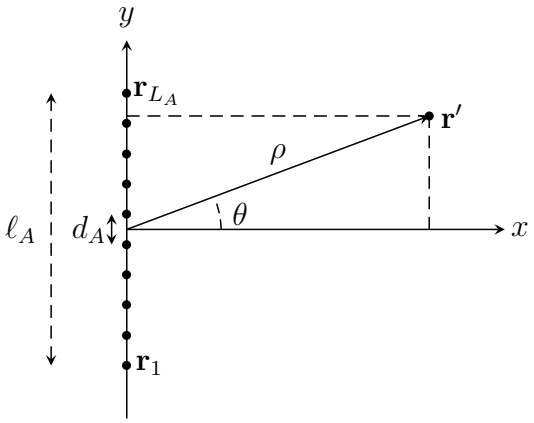


Figure 3.2: The ULA and the coordinate systems of the audio scene.

Source Radiation Pattern As previously described in section 2.2, the radiation pattern $D(\theta, \omega) \in [0, 1]$ describes the intensity of the sound field emitted toward direction θ at frequency ω. We define the *orientation* as the direction ϕ at which the radiation pattern is maximum:

$$\phi(\omega) = \arg\max_{\theta} \quad D(\theta, \omega). \tag{3.2}$$

Source Signal A point source emits a signal $s(t)$ as a function of time t; the source is free to move in the space, and its signal can vary in time and frequency. In order to track these changes we adopt the Short Time Fourier Transform (STFT), which decomposes the signal into possibly overlapped frames, where the signal is assumed to be stationary, and implements the Discrete Fourier Transform over K sub-bands.

The starting point of the analysis stage is the array signal $\mathbf{u}_A(t) = [u_1(t), \ldots, u_{L_A}(t)]^T$, along with its STFT $\mathbf{U}_A(t_\tau, \omega_k)$ at the τ-th time frame and k-th frequency bin:

$$U_l(t_\tau, \omega_k) = \int_{-\infty}^{\infty} u_l(t)\psi(t - t_\tau)e^{-j\omega_k t}dt, \quad l = 1, \ldots, L_A, \tag{3.3}$$

where $\psi(t)$ is the window that selects the time frames.
The array signal can be represented as:

$$\mathbf{U}_A(t_\tau, \omega_k) = (\mathbf{D}_A(\omega_k) \circ \mathbf{g}_A(t_\tau, \omega_k))\, S(t_\tau, \omega_k) + \mathbf{e}(t_\tau, \omega_k), \tag{3.4}$$

where $\circ$ denotes the Hadamard product, $\mathbf{D}_A(\omega_k) \in \mathbb{R}^{L_A \times 1}$ collects the radiation pattern value visible by the array:

$$\mathbf{D}_A(\omega_k) = [D(\theta_1), \ldots, D(\theta_{L_A})]^T, \tag{3.5}$$

$\mathbf{g}_A(t_\tau, \omega_k)$ collects the acoustic transfer function values from the source to the array, $S(t_\tau, \omega_k) \in \mathbb{C}$ is the STFT coefficient of the source signal $s(t)$ and $\mathbf{e}(t_\tau, \omega_k)$ is a spatially white noise.

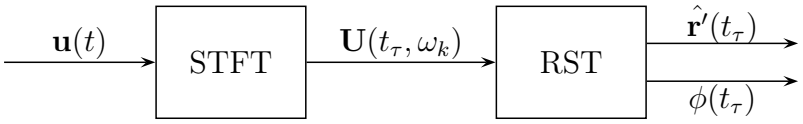


Figure 3.3: Block diagram of the localization.

At the end of the analysis stage, the source signal $s(t)$ can be reconstructed through the Inverse Short Time Fourier Transform (ISTFT):

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(t_\tau, \omega_k) e^{j\omega t} dt_\tau d\omega_k. \tag{3.6}$$

3.1.1 Source Localization

Let $\mathbf{u}_A(t)$ be the microphone array signal and $\mathbf{U}_A(t_\tau, \omega_k)$ its STFT coefficients vector at the τ-th time frame and k-th frequency bin. In the following, the time dependency is neglected for the sake of simplicity; adopting the STFT paradigm, the analysis can be performed either at each time frame or by taking an average over a desired number of time frames, depending on the desired temporal resolution for tracking the source motion.

As shown in section 2.1, we can compute the Ray Space Transform of the array signal, obtaining the coefficient matrix $\mathbf{Y}(\omega_k) \in \mathbb{C}^{I \times W}$

$$\mathbf{Y}(\omega_k) = \mathbf{\Psi}(\omega_k)^H \mathbf{U}_A(\omega_k). \tag{3.7}$$

By taking the magnitude of the ray space coefficients, the wideband magnitude estimate is defined as

$$\mathcal{Y} = \left(\prod_{k=0}^{K/2} E\left\{ |\mathbf{Y}(\omega_k)| \right\} \right)^{2/K}, \quad \mathcal{Y} \in \mathbb{R}^{I \times W}. \tag{3.8}$$

From $\mathcal{Y}$ one peak is identified for each row $\mathcal{Y}_{(i,:)}$ at values $\tilde{m}_i$ and amplitude w_i. The position of the acoustic source is estimated through a weighted least-squares regression [32] on the peaks values and positions. In particular, the founded peaks are collected in the vectors $\tilde{\mathbf{m}} = [\tilde{m}_0, \ldots, \tilde{m}_{I-1}]^T$; then we introduce the matrices $\tilde{\mathbf{M}} = [-\tilde{\mathbf{m}}, \mathbf{1}]$ and $\mathbf{W} = \mathrm{diag}\,(w_0, \ldots, w_{I-1})$; finally we build the vector $\mathbf{q} = [q_0, \ldots, q_{I-1}]^T$ that collects the y-coordinate of the centers of the spatial windows.

The source position can then be estimated as

$$\hat{\mathbf{r}}' = \left(\tilde{\mathbf{M}}^T \mathbf{W} \tilde{\mathbf{M}} \right)^{-1} \tilde{\mathbf{M}}^T \mathbf{W} \mathbf{q}. \tag{3.9}$$

3.1.2 Orientation and Radiation Pattern Extraction

In this section we present the methodology for estimating the orientation and the radiation pattern from the array signal, and the subsequent decomposition in circular harmonics for interpolating the radiation pattern

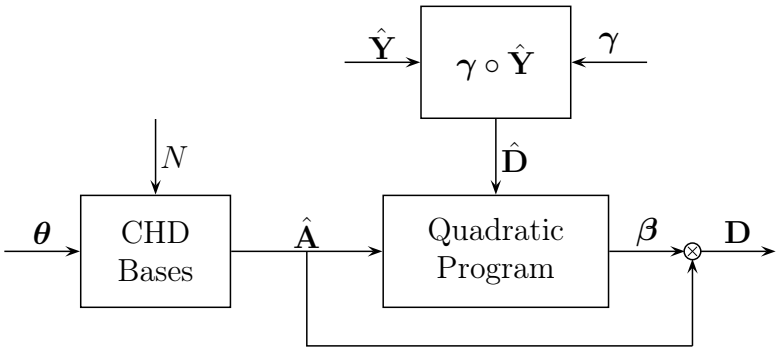


Figure 3.4: Block diagram of the radiation pattern extraction.

over the whole directional domain. Since the radiation pattern is a function of the frequency ω also its modelling is frequency-dependent. For the sake of semplicity, without loss of generality, we neclegt the frequency dependency in the following description. Figure 3.4 shows the processing for the radiation pattern estimation and modelling through the Circular Harmonics Decomposition.

3.1.2.1 Orientation Estimation

Let $\boldsymbol{\rho} \in \mathbb{R}^{I \times 1}$ be the vector that collects the distances between the source position and each center of the spatial windows:

$$\{\boldsymbol{\rho}\}_i = \|\mathbf{q}_i - \hat{\mathbf{r}'}\|, \tag{3.10}$$

and $\hat{\mathbf{m}}, \boldsymbol{\theta} \in \mathbb{R}^{I \times 1}$ vectors that collect, respectively, the RS linear pattern and direction from the source position and each center of the spatial windows:

$$\begin{aligned} \{\hat{\mathbf{m}}\}_i &= \frac{\hat{y'} - q_i}{\hat{x'}} \\ \{\boldsymbol{\theta}\}_i &= \arctan \{\hat{\mathbf{m}}\}_i \end{aligned} \tag{3.11}$$

Along with the geometry we include an acoustic transfer function $\mathbf{g}(\omega) \in \mathbb{C}^{I \times 1}$ from any couple of points in the space. In particular, we choose to adopt the Green function of section 1.1.2:

$$\{\mathbf{g}\}_i (\omega) = \frac{e^{-j\frac{\omega}{c}\rho_i}}{4\pi\rho_i}. \tag{3.12}$$

The acoustic transfer function (3.12) is adopted in order to compensate the ray space coefficients matrix for the distance by constructing a new matrix $\hat{\mathbf{Y}}(\omega_k)$ as follows:

$$\left\{\hat{\mathbf{Y}}\right\}_{(i,:)} = \frac{\{\mathbf{Y}\}_{(i,:)}}{|\{\mathbf{g}\}_i|} = 4\pi\rho_i \{\mathbf{Y}\}_{(i,:)}, \tag{3.13}$$

where $(i, :)$ denotes the i-th row of the matrix.

Finally, under the assumption that the source is oriented torwards the array, i.e. the radiation pattern maximum is visible, the orientation can be estimated from matrix in (3.13). Indeed, given the definition (3.2), we can relate the maximum absolute value of the compensated coefficients matrix to the orientation:

$$\begin{aligned} \mu &= \arg\max_{m} \quad \hat{\mathbf{Y}} \\ \hat{\phi} &= \arctan(\mu). \end{aligned} \tag{3.14}$$

3.1.2.2 Radiation Pattern Estimation

From the localization step we inherit the matrix $\hat{\mathbf{Y}}$ (3.13) that encodes the ray space coefficients (m, q) compensated for the distance from the source to the centers of the spatial windows. Given the array signal model (3.4), we can state that the matrix $\hat{\mathbf{Y}}$ includes in its coefficients also the information about the *pattern-times-signal* $\mathbf{D}_A S$. In particular, the radiation pattern is estimated by reading the absolute value of $\hat{\mathbf{Y}}$ along the linear pattern that emerges from the localization:

$$\left\{\hat{\mathbf{D}}\right\}_i = \{\boldsymbol{\gamma}\}_i \cdot \left\{|\hat{\mathbf{Y}}|\right\}_{(q_i, \tilde{m}_i)}, \tag{3.15}$$

where $\boldsymbol{\gamma}$ compensates for the gaussian window of the Ray Space Transform (2.10):

$$\{\boldsymbol{\gamma}\}_i = \left(d_A \sum_{l=1}^{L_A} e^{-\pi (y_l - q_i)^2 / \sigma^2} \right)^{-1}. \tag{3.16}$$

3.1.2.3 Circular Harmonics Decomposition

Once the radiation pattern has been estimated, we would perform the circular harmonics decomposition of section 3.1.2.3. In particular, we state that

$$D(\theta) = \sum_{n=0}^{N-1} b_n \cos\left(n\left(\theta - \phi\right)\right), \tag{3.17}$$

where N is the order of the decomposition and b_n the n-th order coefficient. By Exploiting simple trigonometric relations

$$\cos(n(\theta - \phi)) = \cos(n\theta - n\phi) = \cos(n\theta)\cos(n\phi) + \sin(n\theta)\sin(n\phi)$$

and defining

$$\begin{aligned} \beta_n^{(c)} &= b_n \cos n\phi \\ \beta_n^{(s)} &= b_n \sin n\phi \end{aligned} \tag{3.18}$$

the decomposition becomes

$$D(\theta) = \sum_{n=0}^{N-1} \left[\beta_n^{(c)} \cos n\theta + \beta_n^{(s)} \sin n\theta\right] \tag{3.19}$$

The Circular Harmonics Decomposition can be conveniently written in matrix form by defining a matrix $\hat{\mathbf{A}} \in \mathbb{R}^{I \times 2N}$ collecting the cosine and sine functions:

$$\hat{\mathbf{A}} = \begin{bmatrix} \cos 0 & \dots & \cos(N-1)\theta_0 & \sin 0 & \dots & \sin(N-1)\theta_0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos 0 & \dots & \cos(N-1)\theta_{I-1} & \sin 0 & \dots & \sin(N-1)\theta_{I-1} \end{bmatrix} \tag{3.20}$$

and a vector $\boldsymbol{\beta} \in \mathbb{R}^{2N \times 1}$ that collects the coefficients in (3.18):

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0^{(c)} \dots \beta_{N-1}^{(c)} & \beta_0^{(s)} \dots \beta_{N-1}^{(s)} \end{bmatrix}^T \tag{3.21}$$

yielding to the radiation pattern evaluated in the range of directions $\boldsymbol{\theta}$ (3.11):

$$\mathbf{D} = \hat{\mathbf{A}}\boldsymbol{\beta}. \tag{3.22}$$

Adopting this model, the vector of CHD coefficients $\boldsymbol{\beta}$ can be estimated from the radiation pattern estimation in (3.15) using a linearly constrained minimization problem:

$$\begin{aligned} \arg\min_{\boldsymbol{\beta}} \quad & \|\hat{\mathbf{D}} - \hat{\mathbf{A}}\boldsymbol{\beta}\|^2 \\ \text{subject to} \quad & \mathbf{A}\boldsymbol{\beta} \geq 0 \end{aligned} \tag{3.23}$$

We have introduced a matrix $\mathbf{A} \in \mathbb{R}^{N_\theta \times 2N}$ constructed similarly to $\hat{\mathbf{A}}$ but for the number of rows N_θ, which is the number of directions resulting from a uniform sampling of the direction domain with sampling interval $\bar{\theta}$:

$$N_\theta = 2\pi/\bar{\theta}. \tag{3.24}$$

Note that in this problem (3.23) the radiation pattern is estimated on a limited set of directions (encoded by the matrix $\hat{\mathbf{A}}$), while the constraints must be respected $\forall\theta \in [0, 2\pi]$.

By exploiting some algebraic computations, one can obtain

$$\|\hat{\mathbf{D}} - \hat{\mathbf{A}}\boldsymbol{\beta}\|^2 = \hat{\mathbf{D}}^T\hat{\mathbf{D}} - 2\hat{\mathbf{D}}^T\hat{\mathbf{A}}\boldsymbol{\beta} + \boldsymbol{\beta}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\boldsymbol{\beta}, \tag{3.25}$$

where the first term is independent on the minimization variable and thus it acts as a bias, the second term is linear and the last term is quadratic. This yields to a more convenient formulation in the form of a standard quadratic programming problem [33]:

$$\begin{aligned} \arg\min_{\boldsymbol{\beta}} \quad & \frac{1}{2}\boldsymbol{\beta}^T\mathbf{Q}\boldsymbol{\beta} + \mathbf{c}^T\boldsymbol{\beta} \\ \text{subject to} \quad & \tilde{\mathbf{A}}\boldsymbol{\beta} \leq \mathbf{b}, \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} \mathbf{c} &= -2\hat{\mathbf{A}}^T\hat{\mathbf{D}} & \mathbf{Q} &= 2\hat{\mathbf{A}}^T\hat{\mathbf{A}} \\ \tilde{\mathbf{A}} &= -\mathbf{A} & \mathbf{b} &= [\mathbf{0}] \quad \in \mathbb{Z}^{N_\theta \times 1} \end{aligned} \tag{3.27}$$

Furthermore, we can reasonably assume that the source is oriented towards the array, i.e. the array is able to sense the maximum of the radiation pattern 3.2. As a consequence, two new constraints are added to the minimization problem 3.23:

$$\begin{aligned} &\left.\frac{\partial D(\theta)}{\partial \theta}\right|_{\theta=\hat{\phi}} = 0 \\ &D(\hat{\phi}) \geq D(\theta) \quad \forall \theta \in [0, 2\pi]. \end{aligned} \tag{3.28}$$

We can compute the first derivative in the circular harmonics space as follows:

$$\begin{aligned} &\left.\frac{\partial D(\theta)}{\partial \theta}\right|_{\theta=\hat{\phi}} = \sum_{n=0}^{N-1} -\beta_n^{(c)} n \sin(n\hat{\phi}) + \beta_n^{(s)} n \cos(n\hat{\phi}) = \bar{\mathbf{a}}\boldsymbol{\beta}, \\ &\bar{\mathbf{a}} = \left[0, \ldots, -(N-1)\sin\left((N-1)\hat{\phi}\right) \quad 0, \ldots, (N-1)\cos\left((N-1)\hat{\phi}\right)\right] \end{aligned} \tag{3.29}$$

where $\bar{\mathbf{a}} \in \mathbb{R}^{1\times 2N}$ collects the first derivative of the CH basis functions computed at $\hat{\phi}$. By introducing a new matrix $\bar{\mathbf{A}} \in \mathbb{R}^{N_\theta \times 2N}$ constituted by the same row vector containing the CH basis functions computed at $\hat{\phi}$ and stacked N_θ times

$$\left\{\bar{\mathbf{A}}\right\}_{row} = \left[1, \ldots, \cos\left((N-1)\hat{\phi}\right) \quad 0, \ldots, \sin\left((N-1)\hat{\phi}\right)\right], \tag{3.30}$$

the quadratic programming problem becomes

$$\begin{aligned} \arg\min_{\boldsymbol{\beta}} \quad & \frac{1}{2}\boldsymbol{\beta}^T \mathbf{Q}\boldsymbol{\beta} + \mathbf{c}^T\boldsymbol{\beta} \\ \text{subject to} \quad & \tilde{\mathbf{A}}\boldsymbol{\beta} \leq \mathbf{b} \\ & \bar{\mathbf{a}}\boldsymbol{\beta} = 0, \end{aligned} \tag{3.31}$$

having

$$\begin{aligned} \mathbf{c} &= -2\hat{\mathbf{A}}^T\hat{\mathbf{D}} & \mathbf{Q} &= 2\hat{\mathbf{A}}^T\hat{\mathbf{A}} \\ \tilde{\mathbf{A}} &= \begin{bmatrix} -\mathbf{A} \\ \mathbf{A} - \bar{\mathbf{A}} \end{bmatrix} & \mathbf{b} &= [\mathbf{0}] \quad \in \mathbb{Z}^{2N_\theta \times 1}. \end{aligned} \tag{3.32}$$

3.1.3 Signal Extraction

So far, the sound scene has been processed entirely in the ray space: the localization through a weighted least squares problem, the orientation through a peak extraction. The results have been forwarded to the quadratic programming problem (3.23) for radiation pattern estimation and decomposition. The descriptors achieved so far can be now arranged in order to extract the source signal directly from the array signal $\mathbf{U}_A$. We need to compute the radiation pattern $\mathbf{D}_A$ seen from each l-th microphone, along with the acoustic transfer function $\mathbf{g}_A$ from the estimated source position $\hat{\mathbf{r}}'$ to the sensor position $\mathbf{r}_l = [x_l, y_l]^T$. First, a new ma-

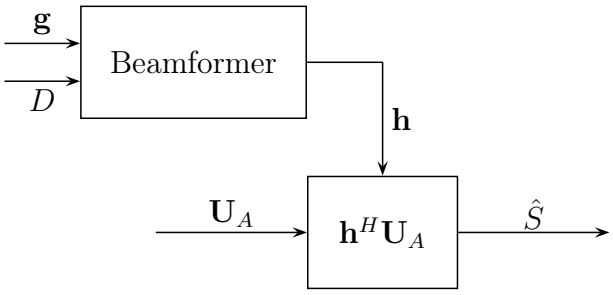


Figure 3.5: Block diagram of the signal extraction.

trix of CH basis functions $\hat{\mathbf{A}}_A \in \mathbb{R}^{L_A \times 2N}$ must be computed for those directions $\boldsymbol{\theta}_A = [\theta_1, \ldots, \theta_{L_A}]^T$ from the source towards each microphone:

$$\hat{\mathbf{A}}_A = \begin{bmatrix} \cos 0 & \ldots & \cos(N-1)\theta_1 & \sin 0 & \ldots & \sin(N-1)\theta_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos 0 & \ldots & \cos(N-1)\theta_{L_A} & \sin 0 & \ldots & \sin(N-1)\theta_{L_A} \end{bmatrix}, \tag{3.33}$$

yielding to

$$\mathbf{D}_A = \hat{\mathbf{A}}_A \boldsymbol{\beta}, \tag{3.34}$$

Recalling the array signal model of (3.4):

$$\mathbf{U}_A = (\mathbf{D}_A \circ \mathbf{g}_A)\, S + \mathbf{e}, \tag{3.35}$$

since we are interested in extracting the source signal S, we aim at designing a beamformer of the form

$$\mathbf{h}^H \mathbf{U}_A = \mathbf{h}^H (\mathbf{D}_A \circ \mathbf{g}_A)\, S + \mathbf{h}^H \mathbf{e}. \tag{3.36}$$

Thus the minimization problem is similar to the Delay-and-Sum beamformer introduced in section 1.3.2.2:

$$\begin{aligned} \arg\min_{\mathbf{h}} \quad & \mathbf{h}^H \mathbf{h} \\ \text{subject to} \quad & \mathbf{h}^H (\mathbf{D}_A \circ \mathbf{g}_A) = 1, \end{aligned}$$

yielding to the *informed* beamformer:

$$\mathbf{h} = \frac{\mathbf{D}_A \circ \mathbf{g}_A}{\|\mathbf{D}_A \circ \mathbf{g}_A\|} \tag{3.37}$$

Finally, the source signal is extracted through a simple vector multiplication:

$$\hat{S} = \mathbf{h}^H \mathbf{U}_A \tag{3.38}$$

and, once all the time frames and frequency bins have been analysed, the original signal can be reconstructed via the Inverse Short Time Fourier Transform:

$$\hat{s}(t) = ISTFT\left\{\hat{S}(t_\tau, \omega_k)\right\}. \tag{3.39}$$

3.2 Scene Synthesis

This section describes the spatial sound scene manipulation and synthesis, based on the data extracted in the analysis stage. The manipulation refers to the source translation and the rotation with respect to its axis. Figure 3.6 helps visualizing these two operations. In the following we show the synthesis steps assuming to have a uniform linear array of L_S elements, with a interdistance d_S.

3.2.1 Geometry Synthesis

Being $\mathbf{r'}_S(t)$ the desired source position, the geometric parameters and the acoustic transfer function are immediately computed as follows:

$$\{\boldsymbol{\theta}_S(t)\}_l = \arctan \frac{y_l - y'_S(t)}{x_l - x'_S(t)}, \quad l = 1, \dots, L_S, \tag{3.40}$$

$$\{\boldsymbol{\rho}_S(t)\}_l = \sqrt{(y_l - y'_S(t))^2 + (x_l - x'_S(t))^2}, \quad l = 1, \dots, L_S, \tag{3.41}$$

$$\{\boldsymbol{g}_S(t,\omega)\}_l = \frac{e^{-j\frac{\omega}{c}\rho_l(t)}}{4\pi\rho_l(t)}, \quad l = 1, \dots, L_S. \tag{3.42}$$

Note that the acoustic transfer function can be either equal to the analysis stage (i.e. a Green function) or different, if a new model fits better the synthesis conditions.

3.2.2 Radiation Pattern Synthesis

Being ϕ_S the desired orientation, the new set of directions is

$$\tilde{\boldsymbol{\theta}}_S(t) = \Delta\phi(t) + \boldsymbol{\theta}_S(t), \tag{3.43}$$

where

$$\Delta\phi(t) = \phi_S(t) - \hat{\phi}(t). \tag{3.44}$$

A new matrix of CHD basis functions is $\mathbf{A}_S(t)$ is built as (3.20)

$$\hat{\mathbf{A}}_S = \begin{bmatrix} \cos 0 & \dots & \cos(N-1)\tilde{\theta}_1 & \sin 0 & \dots & \sin(N-1)\tilde{\theta}_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos 0 & \dots & \cos(N-1)\tilde{\theta}_{L_S} & \sin 0 & \dots & \sin(N-1)\tilde{\theta}_{L_S} \end{bmatrix}. \tag{3.45}$$

Thanks to the coefficients vector $\boldsymbol{\beta}(\omega)$ estimated by the quadratic programming problem (3.26), we can synthesize the radiation pattern:

$$\mathbf{D}_S(t,\omega) = \mathbf{A}_S(t)\boldsymbol{\beta}(\omega). \tag{3.46}$$

Alternatively, we can adopt the simple CHD model of equation (3.19) and compute the cosine functions coefficients by exploiting the trigonometric relations of (3.18):

$$b_n = \sqrt{{\beta_n^{(c)}}^2 + {\beta_n^{(s)}}^2}, \quad \forall n \in [0, N-1]. \tag{3.47}$$

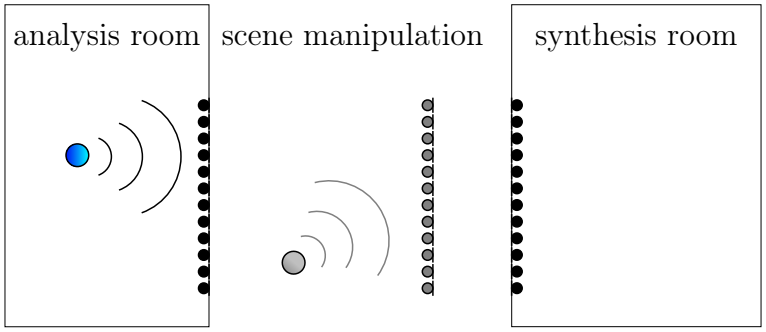


Figure 3.6: The acoustic curtain paradigm.

Finally, by introducing the CHD basis matrix

$$\hat{\mathbf{A}}_S = \begin{bmatrix} \cos 0 & \dots & \cos(N-1)\tilde{\theta}_1 \\ \vdots & \ddots & \vdots \\ \cos 0 & \dots & \cos(N-1)\tilde{\theta}_{L_S} \end{bmatrix} \tag{3.48}$$

the radiation pattern $\mathbf{D}_S$ is computed by

$$\mathbf{D}_S(t,\omega) = \hat{\mathbf{A}}_S(t)\mathbf{b}(\omega), \quad \mathbf{b} = [b_0, \dots, b_{N-1}]^T. \tag{3.49}$$

3.2.3 Signal Synthesis

Adopting the time-frequency processing paradigm, we can compute the array signal $\mathbf{U}_S(t_\tau, \omega_k) \in \mathbb{R}^{L_S \times 1}$ for each τ-th time frame and k-th frequency bin. The STFT parameters must be inherited from the analysis stage in order to guarantee the consistency of the synthesized signals.

The signal model is:

$$\mathbf{U}_S(t_\tau, \omega_k) = \mathbf{g}_S^{j\alpha}(t_\tau, \omega_k)\mathbf{D}_S(\omega_k)\hat{S}(t_\tau, \omega_k), \tag{3.50}$$

where the parameter α controls the sign of the complex exponential argument. If the rendering area is the same of the analysis area (i.e. for active room control [34], [35], [36]), the sign is the same of the analysis transfer function (i.e. a minus), while if the rendering system is designed as an *acoustic curtain* [6] for continuing the wave propagation, the sign has to change (i.e. a plus):

$$\alpha = \begin{cases} 1, & \text{if rendering area} = \text{analysis area} \\ -1, & \text{if rendering area} \neq \text{analysis area} \end{cases} \tag{3.51}$$

Finally, the array signal $\mathbf{u}_S(t)$ is reconstructed by the Inverse Short Time Fourier Transform (3.6):

$$\mathbf{u}_S(t) = ISTFT\left\{\mathbf{U}_S(t_\tau, \omega_k)\right\} \tag{3.52}$$

3.3 Conclusive Remarks

In this chapter we propose a synthesis-by-analysis framework for sound scene manipulation. The localization is performed in the (m, q) ray space of [4] that enables the usage of linear pattern analysis techniques on the plenacoustic image. The radiation pattern is estimated by reading the value of the ray space transform (compensated for the distance between the source and the array) along the linear pattern that emerges from the localization stage; then the radiation pattern is decomposed in the Circular Harmonics domain and its parameters estimated by a linearly constrained quadratic programming problem. Finally, the source signal is extracted through a beamforming operation, which employs the radiation pattern and the acoustic transfer function.

The information emerged from the analysis stage is provided to the rendering system, allowing the user to arbitrarily choose the position and the orientation of the virtual source. In this proposal, we assume to use the same STFT parameters for both the analysis and synthesis stages; however, the agnosticism of the rendering system could be improved by interpolating the radiation pattern along the time axis. As a matter of facts, the directivity pattern is strongly related to the physical characteristics of the source, thus its time-invariance can be assumed for most multimedia applications.

4
Simulations and Tests

In this chapter we validate the sound scene manipulation methodology proposed in this thesis. Firstly we will briefly describe the metrics used for the evaluation. Then we will simulate a real acoustic scenario using a monopole source emitting white noise signals. This kind of analysis is particularly useful since, given the wide-band nature of the emitted signals, we can evaluate the performance of our approach in a wide range of frequencies. Finally, we show the behavior of the system on a real applicative scenario, in order to obtain the metrics of interest, using microphones and a loudspeaker arranged in different scene geometries.

4.1 Evaluation Metrics

Localization In order to validate the localization approach proposed in section 3.1.1, we measure the performance with the localization error, defined as the distance between the actual source position $\mathbf{r}'$ and the estimated one:

$$\epsilon = \|\mathbf{r}' - \hat{\mathbf{r}'}\|, \tag{4.1}$$

where $\hat{\mathbf{r}'}$ is obtained through (3.9) applied to normalized versions of the RS coefficients matrices in order to avoid numerical problems.

Orientation As for the localization, we measure the performance with the orientation error, defined as the distance between the actual source orientation ϕ and the $\hat{\phi}$ estimated from (3.14):

$$\xi = |\phi - \hat{\phi}|. \tag{4.2}$$

We will consider the expected value of both the localization and orientation errors. More precisely, we will compute the sample mean over J realizations:

$$\begin{aligned} E\{\epsilon\} &= \frac{1}{J}\sum_{j=0}^{J-1} \|\mathbf{r}' - \hat{\mathbf{r}}'^{(j)}\|, \\ E\{\xi\} &= \frac{1}{J}\sum_{j=0}^{J-1} |\phi - \hat{\phi}^{(j)}|. \end{aligned} \tag{4.3}$$

Synthetised Soundfield In order to validate the synthesized pressure field $p_S(y,\omega)$ against the pressure field as if it would be generated by a virtual source, denoted by $p_V(y,\omega)$, we compare the microphone signals $u_l(t), l = 1,\dots,L$ of the desired sound field to the synthesized microphone signals $\hat{u}_l(t)$. We compute the average over the whole microphone array of the ratio of the error energy and the actual signal energy for each microphone.

$$NMSE = 10\log_{10}\left(\frac{1}{L}\sum_{l=1}^{L}\frac{\sum_t |u_l(t)-\hat{u}_l(t)|^2}{\sum_t |u_l(t)|^2}\right) \tag{4.4}$$

4.2 Simulations

We have considered a single point-like source emitting a white noise signal. The acoustic environment is assumed not to produce any reverberation, thus the free-field condition is respected.

Scene Geometry As depicted in picture 4.1, the acoustic scene is composed by

- a point source placed in $\mathbf{r}' = [1;0]^T$ m;
- the source radiation pattern is a cardiod with orientation $\phi = \pi$, i.e. towards the center of the array. In particular, we have adopted the model (3.19) with $\mathbf{b} = [1/2;1/2]^T$;
- the sensor array is composed by $L = 16$ omnidirectional microphones with inter-distance $d = 10$ cm, displaced along the y axis.

Ray Space Transform For the computation of the Ray Space Transform matrix $\mathbf{\Psi}$ (2.15) we have adopted the parameters reported by table 4.1.

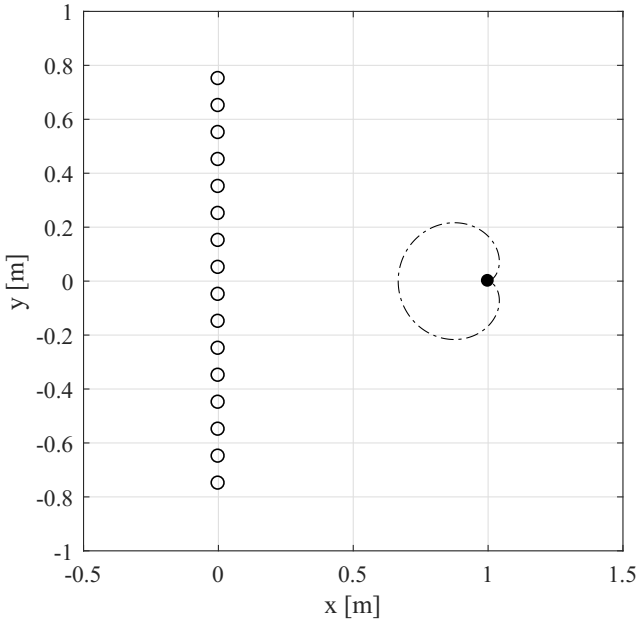


Figure 4.1: The simulated scene geometry.

q axis sampling interval	$\bar{q} = d/2 = 5$ cm
Total spatial windows	$I = 32$
m axis sampling interval	$\bar{m} \approx 0.0028$
Total visible directions	$W = 350$
Angular aperture	$m = \pm 4 \quad (\theta \approx \pm 76°)$
Gaussian windows standard deviation	$\sigma = 30$ cm

Table 4.1: Ray Space Transform parameters.

Circular Harmonics Decomposition The order of the CH decomposition is chosen with respect to the frequency. In fact, we can reasonably assume that at lower frequencies the physical characteristics of the sources results in a smoother radiation pattern, such as omni-directional ($N = 1$) or cardioid-like ($N = 2$). At higher frequencies, instead, most of the acoustic sources exhibit preferences for certain directions, therefore an higher order for the CHD could fit better the received radiation pattern estimates. More precisely, we have arbitrarly set the following conditions:

$$N = \begin{cases} 2 & \text{for } f \leq 400 \text{ Hz} \\ 3 & \text{for } 400 \text{ Hz} < f \leq 800 \text{ Hz} \\ 4 & \text{for } 800 \text{ Hz} < f \leq 1.6 \text{ kHz} \\ 5 & \text{for } f > 1.6 \text{ kHz}. \end{cases} \tag{4.5}$$

Short Time Fourier Transform In the simulation we have considered only one time frame, and both the source signal and the additive noise were generated in the frequency domain. Although it is improper to refer to a Short Time version of the Fourier Transform, nevertheless

we address the analysis as if the time dependence would be considered. Table 4.2 reports the STFT parameters. The frequency range is limited by the spatial aliasing of the array, which is defined as the frequency at which the wavelength $\lambda = c/f$ is double the distance between the array elements d:

$$f < \frac{c}{2d} \approx 1700 \text{ Hz} \tag{4.6}$$

Frequency Range [Hz]	$250 \div 1700$
Window Type	Hamming
Window Length [ms]	50
Overlap	75%
Sampling frequency	8 kHz
FFT bins	512

Table 4.2: STFT parameters.

4.2.1 Scene Analysis

In this section we describe the results of the scene analysis stage. In this simulation the time variable is neglected for the sake of semplicity. The source signal S is propagated towards the array through a Green function which exploits the scene geometry. As described in the previous chapter 3, the matrix $\mathbf{U} \in \mathbb{C}^{L \times K/2}$ collects in its rows the spectra of the microphone signals, as done in (3.4):

$$\mathbf{U} = (\mathbf{D} \circ \mathbf{g})S + \mathbf{e} \tag{4.7}$$

where $\mathbf{e}$ is a white noise added to the array signal. The processing algorithm is described by figure 3.3. The results are averaged over $J = 10$ realizations.

Figure 4.2 shows how the localization and orientation errors vary with respect to four different SNR values, which are computed by taking the central microphone as reference. The localization is robust to different levels of SNR. The orientation error, instead, goes to zero as the noise level decreases. The values of these errors are negligible with respect to the human hearing capabilities [37].

4.2.2 Scene Synthesis

In this section we describe how the sound scene manipulation methodology can be performed in the simulated scenario. In order to better understand the synthesis capabilities of the system, three different simulations have been set up. First, the orientation dependence is tested by synthetizing a pressure field at the microphone array as it would be generated by the analysed source without changing its position $[\theta, \rho] = [0 \text{ rad}, 1 \text{ m}]$. Second, the orientation is fixed to $\phi = \pi$ in order to test the dependence

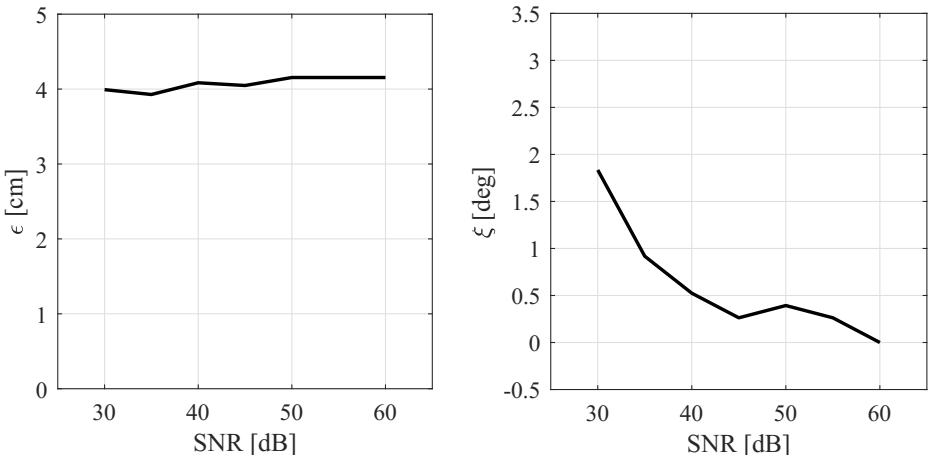


Figure 4.2: The localization error (left) and the orientation error (right) for the simulated scene.

on angular coordinate θ upon fixing $\rho = 1$ m. Third, the angular coordinate is fixed to $\theta = 0$ rad and the system dependence on the distance is tested. We separate the case of a source moving towards the array and the case of a source moving far from the array. The synthetized signal is evaluated with the NMSE presented in section 4.1; the results are provided with respect to the *Signal-to-Noise Ratio* (SNR) values of the microphone array signal. We observe that for higher values of SNR the synthesis behaves better; moreover the NMSE values do not increase so much with respect to the SNR values, meaning that the system seems to be robust to the additive spatially white noise.

Impact of the orientation on the soundfield synthesis The source is rotated about its axis in order to keep the orientation in the viibility range of the array. The evaluation set of orientation is $\Phi = \{\phi|\theta_1 \leq \phi \leq \theta_L\}$, sampled each 5 degrees. Figure 4.3 shows the results. As expected, the system performance degrades when the synthesized source assumes an orientation different from the analysed source orientation. This behavior can be explained by considering that the radiation pattern is extrapolated on the analysis set of directions but synthesized on a different set of directions.

Impact of the angular coordinate on the soundfield synthesis The source moves on a circumference of radius $\rho = 1$ m keeping its original orientation $\phi = \pi$. The evaluation set is $\Theta = \{\theta|y_1 \leq y_S \leq y_L\}$, sampled each 5 degrees. Figure 4.4 shows the results. As expected, the system performance degrades when the synthesis position moves away from the analysis position. This behavior can be explained by considering that the radiation pattern is extrapolated on the analysis set of directions but synthesized on a different set of directions.

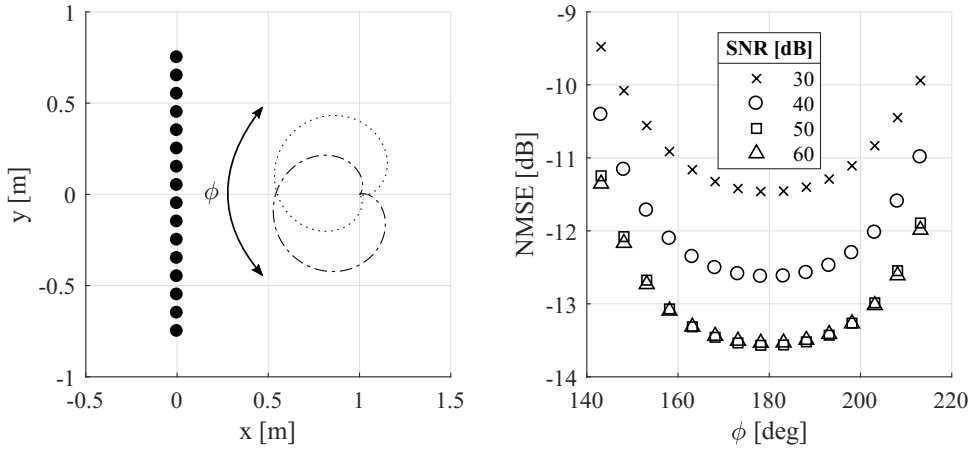


Figure 4.3: NMSE values when manipulating the source orientation.

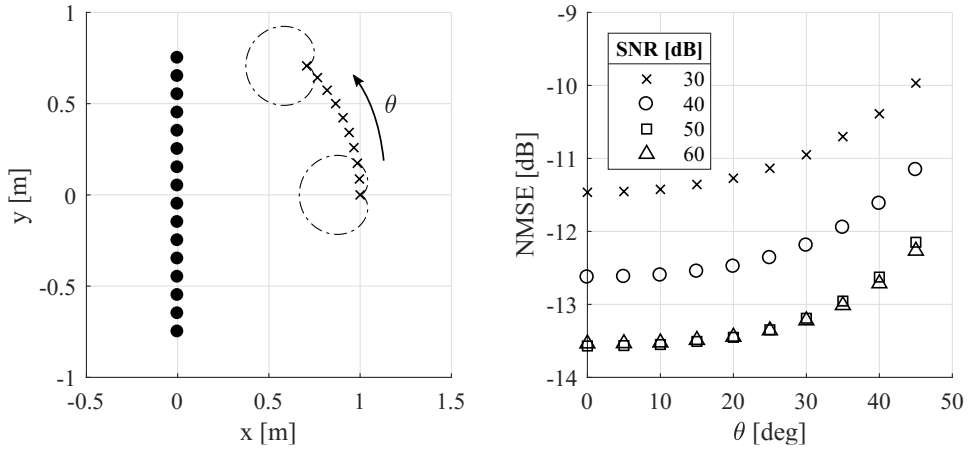


Figure 4.4: NMSE values (right) when manipulating the angular coordinate (left).

Impact of the distance on the soundfield synthesis The source moves on a straight line relying on the x axis, i.e. changing only the distance coordinate ρ. The evaluation set is $P = \{\rho | 1 \leq \rho \leq 3\}$, sampled each 15 cm. Figures 4.5 refers to the analysis position $[1; 0]$ increasing ρ; instead figure 4.6 refers to the analysis position $[3; 0]$ decreasing ρ towards the array. As expected, the system capabilities degrade when the source position is synthesized far away from the analysed position. Firstly, the localization and orientation processes degrades at higher distances. Secondly, the radiation pattern is extrapolated and subsequently synthesized on directions sets of different size. In the first simulation of figure 4.5, the source moves away from the array and thus the directions set is reduced. On the other hand, in simulation of 4.6, the source comes closer to the array and thus the directions set enlarges. Moreover, the NMSE values are higher due to the higher localization error.

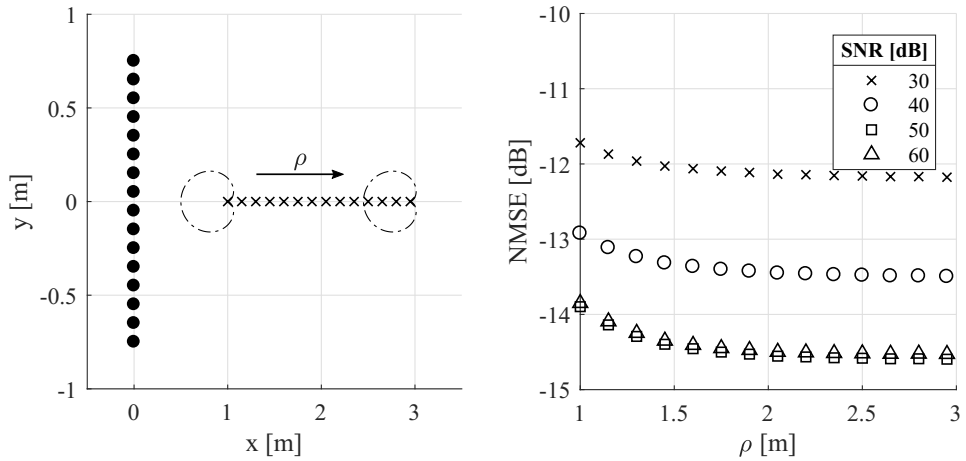


Figure 4.5: NMSE values (right) when moving the source away (left).

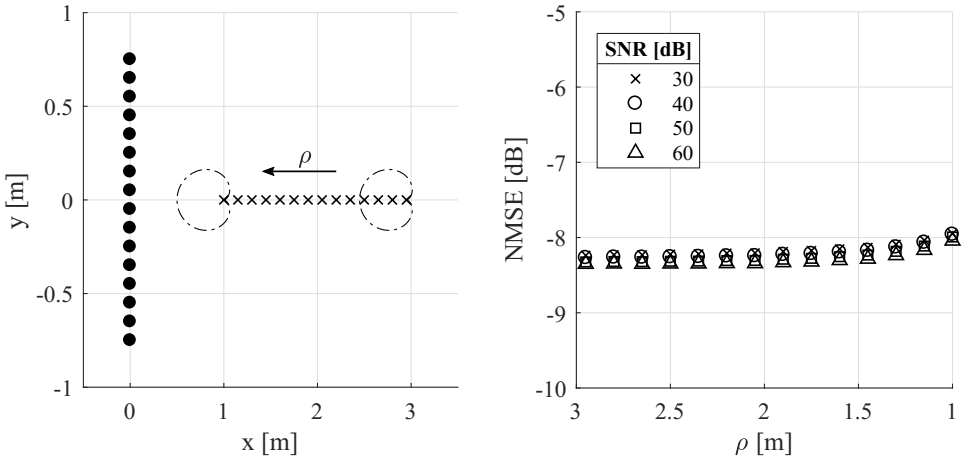


Figure 4.6: NMSE values (right) when moving the source towards the array (left).

4.3 Experiments

In this section we address the experiments we set up in order to test the proposed methodology in a real acoustic scenario. The experiments have been performed in a semi-anechoic with size 4.3×4 m^2 and height 2.6 m. The room is depicted in figure 4.7. With the expression *semi-anechoic room* we refer to a room in which the walls, ceil and floor are covered with sound absorbing material. This treatment is aimed at minimizing reverberation phenomena and allows to get propagation condition closer to free-field. This room as a reverberation time T60 = 50 ms and has been made available for our experiment by the *Sound and Music Computing Lab* of Politecnico di Milano, Como campus.

As depicted in figure 4.7, in this room we have placed a uniform linear array of $L = 16$ microphones with an interdistance $d = 10$ cm. The microphones composing the array are *Beyerdynamic MM1* whose specifications are reported in appendix A.

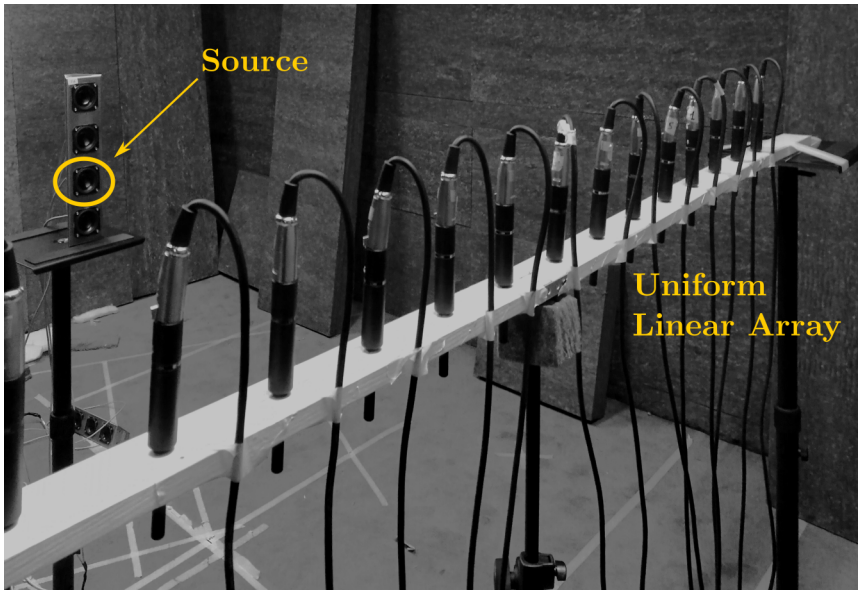


Figure 4.7: The loudspeaker and the microphone array in the experimental setup.

The microphone signals are preamplified, converted and recorded on a Digital Audio Workstation (DAW). Table 4.3 reports the acquisition machinery of the Sound and Music Computing Lab. The microphone

Microphones	Beyerdynamic MM1
Preamplifier	Focusrite Platinum Le Octopre
Loudspeaker	B&C passive speaker
Power amplifier	B&C power amplifier
ADC/DAC	Lynx Aurora 16

Table 4.3: The audio equipment of the experiment.

signals are then processed in MATLAB as done in simulations.

4.3.1 Scene Analysis

The acoustic scene we set up for the analysis stage is similar to the simulated scenario. The loudspeaker is placed in front of the microphone array at coordinates $[1; 0]$ m, oriented to the centre of the array with $\phi = \pi$ and it emits a white noise $s(t)$ for 15 seconds. The recorded signals are processed in order to extract the Short Time Fourier Transform with parameters of table 4.2. The position $\hat{\mathbf{r}}'$, orientation $\hat{\phi}$, the radiation pattern coefficients $\boldsymbol{\beta}$ and signal $\hat{s}(t)$ of the source are then extracted and made available for the synthesis stage.

4.3.2 Scene Synthesis

As for the computer-aided simulations, we manipulate the analysed sound scene in terms of both desired position and desired orientation. The synthesized sound fields are compared to the desired ones using the NMSE as defined in 4.4. For each experiment under evaluation, we manually set the loudspeaker in the desired position and orientation and acquire the microphone signals. Then, in MATLAB we synthesize the microphone signals based on the analysis results.

Impact of the orientation on the soundfield synthesis Using a turntable, the loudspeaker is rotated about its vertical axis in order to keep the orientation in the visibility range of the array. The evaluation set is $\Phi = \{\phi | 150° \leq \phi \leq 210°\}$, sampled each 7 degrees. Figure 4.8 shows the results. As expected, the NMSE is minimum when the desired

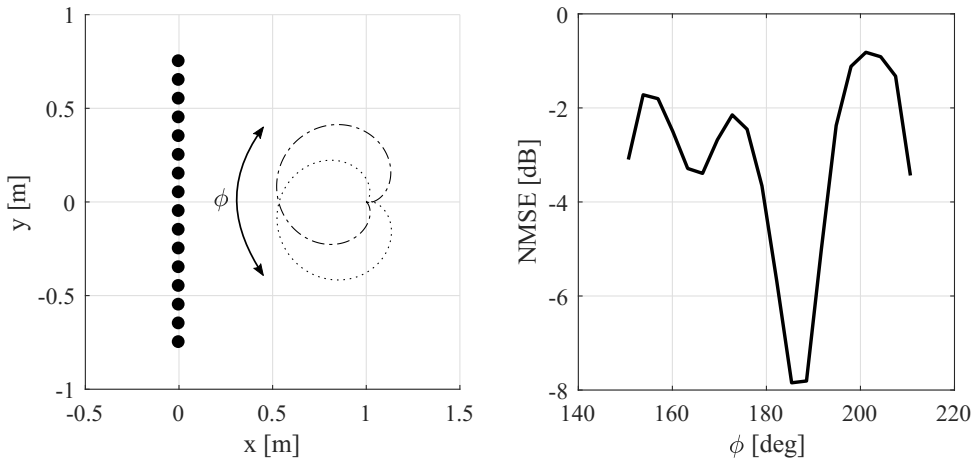


Figure 4.8: NMSE values when manipulating the source orientation.

orientation is similar to the analysed one. When the loudspeaker turns forward the bounds of the array, the evaluation metric grows; this is due to the actual microphone signals, whose energy decreases due to the source orientation, indicating that the error energy is somehow constant over the evaluation settings.

Impact of the angular coordinate on the soundfield synthesis The loudspeaker moves on a circumference or radius $\rho = 1$ m keeping its original orientation. The evaluation set is $\Theta = \{\theta | y_1 \leq y_S \leq y_L\}$, sampled each 10 degrees. Figure 4.9 shows the results. The NMSE curve behaves as expected: performance degrades as the loudspeaker moves away from the position held during the analysis. However, the values do not exhibit an important change, since the source is moving close to some microphones while moving away from others.

Impact of the distance on the soundfield synthesis In the first experiment, the loudspeaker moves on a straight line starting from $[x, y] =$

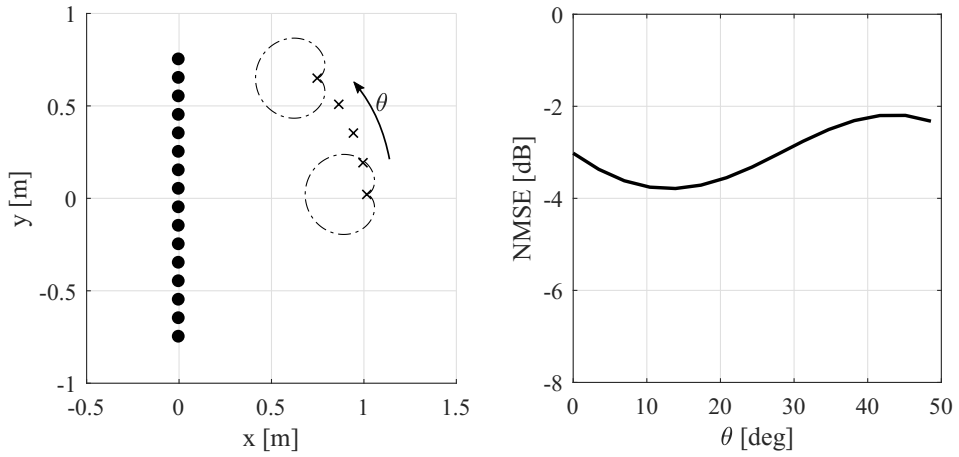


Figure 4.9: NMSE values (right) when manipulating the angular coordinate (left).

$[1, 0]$ (i.e. the analysis position) along the x axis, changing only the distance coordinate ρ. In the second experiment, the analysis is performed at $[3; 0]$ and the source position is synthesized while reducing ρ. The evaluation set is $P = \{\rho | 1 \leq \rho \leq 3\}$, sampled each 15 cm. The system

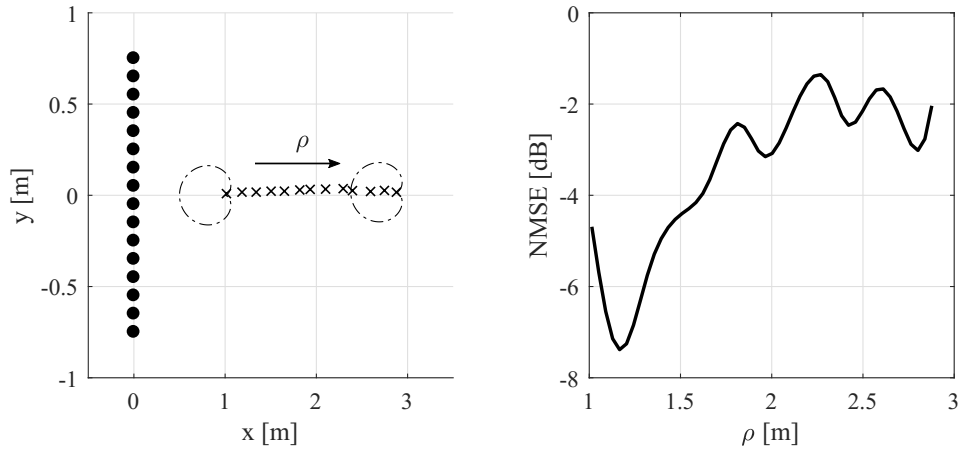


Figure 4.10: NMSE values (right) when moving the source away (left).

capabilities degrade when the source position is synthesized far away from the analysed position. This behavior could be explained by considering that the localization and orientation processes prouce higher errors at higher distances. In addition, the radiation pattern is extrapolated and subsequently synthesized on directions sets of different size. In the first simulation of figure 4.10, the source moves away from the array and thus the directions set is reduced. The energy of the actual microphone signals decreases due to the distance between the source and the array, therefore the NMSE values increases. On the other hand, in simulation of 4.11, the source comes closer to the array and thus the directions set enlarges. The error is expected to increase but it is compensated by the increased energy of the reference signal.

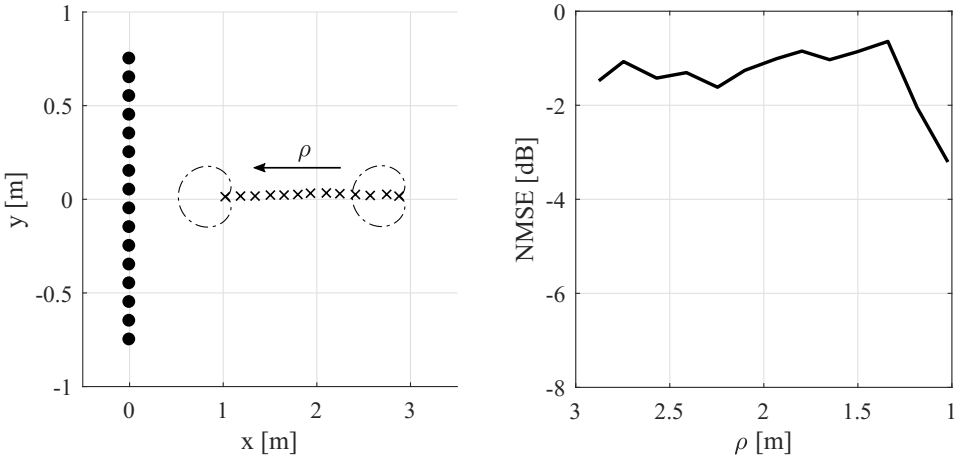


Figure 4.11: NMSE values (right) when moving the source towards the array (left).

4.4 Conclusive Remarks

In this chapter we have evaluated the proposed methodology through simulations and experiments. We started by analysing the sound scene estimate of parameters in a simulated scenario. The localization and orientation errors suggest that the geometrical analysis exhibits robustness to different SNR values applied to the microphone array. The extracted parameters can be easily modified in order to render a desired acoustic scene. We have provided a series of manipulation tests using white noise signals; in particular, we reported the sound field error while rotating and translating the source in space. Then we have tested our method in real scenarios set up in a semi-anechoic chamber. Also in this case our approach is able to retrieve the scene geometry and the source radiation pattern in order to extract the source signal confidently. Overall, the proposed methodology show satisfactory and promising results both in simulation and in the experiments. Moreover, the parametric representation is absolutely format-agnostic, thus the sound scene can be recostructed using different approaches present in the literature.

5
Conclusions and Future Works

This work of thesis proposes a methodology for 2D sound scene manipulation, based on the plenacoustic analysis of a scene and adopting the parametric spatial sound paradigm. The designed system exploits the powerful geometrical representation of the sound field which gives a great insight of the acoustic scene under analysis. This representation has been deeply investigated and very interesting results have come out. Our approach in particular is based on a transformation called Ray Space Transform introduced in [4] that performs a short space-time Fourier transform of the signals captured by a uniform linear array of microphones. This tool maps the array signal onto a domain called ray space. Each point in this domain represents the information of the sound field carried along a ray with a given orientation. In this space, relevant acoustic objects become linear patterns, thus enabling the use of linear pattern analysis techniques. Among the other sound field representations, the adopted one is particularly suitable for the purpose of this thesis, since by acquiring each acoustic ray in a single position in space, the acoustic scenario can be easily reconstructed.

The sound scene parametrization presented in this thesis completely exploits this intuitive representation. In fact, the source is localized through a weighted least squares regression on the peaks of the ray space image. The ray space image is read along the line that is dual of the source position in the geometric space. This information is useful to extract the source radiation pattern visible by the array. Finally, the source signal is extracted by a simple beamforming operation that exploits both the estimated source position and radiation pattern.

The estimate parameters could be easily and intuitively modified in order to manipulate the acoustic scene. In general, the results of the

analysis stage could be provided to any parametric rendering system. We have designed both simulation and experimental sessions in order to validated our methodology. During the simulations, we have tested how our approach behaves when a single acoustic source is present in the sound scene and the microphone signals are affected by a spatially white noise with different SNR values. We observe that the system exhibits robustness to the additive white noise at different SNR values with low performance degradation. Even when important manipulations have been applied to the scene, the whole system performance does not deteriorate significantly. As far as the experiments are concerned, we have tried to reproduce a real scenario in a semi-anechoic chamber. In this situation we have tested our complete system considering different manipulations of the sound scene. The experimental results are consistent with the simulated outcomes. The proposed approach has shown promising results both in simulation and in the experiments.

5.1 Future Works

In this work we have considered a free-field scenario, where a single point source emits a white noise signal. In the following we present three improvements for extend the proposed methodology.

Multiview Approach We would like to generalize to acoustic scenes where more than one sources are present, e.g. a human conversation. In literature many methods for source separation are emerging using a multiview approach [2], [38], [39]. Using multiple Observation Windows could extend the accuracy and robustness of the manipulation system proposed in this thesis, both in terms of localization and extrapolation of the radiation pattern based on a larger number of data.

Reverberant Environment Another possible improvement is related to the extension of the proposed approach to a reverberant environment, i.e. where the free-field cannot be assumed and the sound waves reflect on passive elements (e.g. walls). The microphone array senses the source signal through both direct and reflected propagations. In [2] the authors extend the (m, q) ray space to the projective ray space and showed how to manage multiple primitives in the acoustic scene. With this approach, the reflecting walls can be "recognized" in the analysis stage, thus enabling dereverberation algorithms and allowing the manipulation of passive primitives. Moreover, the portion of radiation pattern visible by the microphone array extends including the Directions of Arrival (DoA) of the sound reflections.

Sparse Circular Harmonics In signal representation, sparsity corresponds to loosely coupled systems, resulting in matrices in which most

of the elements are zero. In section 3.1.2.3 we have set up a linearly constrained quadratic program minimization problem in the Circular Harmonics domain that aims at estimating the set of parameters that best fits the information extracted from the ray space image. The order N of the Circular Harmonics Decomposition is arbitrarly set upon a-priori knowledge about the acoustic source we expect to track. We could adopt a sparse representation in order not to choose arbitrarly the CHD order. Sparse signal processing is emerging as a powerful research field and we expect some interesting inflections in the ray space processing paradigm.

Equipment Specifications

A Beyerdynamic MM1

MM 1

Measurement Microphone

Order #449.350

Features

- Linear frequency response in the diffuse field / under 90°
- Omnidirectional polar pattern
- Calibrated open circuit voltage
- Narrow tubular construction

Applications

The MM 1 is a measurement microphone which has been designed specifically for measuring sound reinforcement and PA-systems. It is designed to work with spectrum analysers for measuring frequency response and sound pressure levels of loud speaker systems. The MM 1 is the ideal microphone for the measurement of audio signals in the research, development, for reverberation testings and other applications.
The narrow tubular construction ensures that the microphone has negligable influence on the sound field so that an increase in sound pressure is avoided with high frequencies. A natural reproduction is achieved due to the linear frequency response.

Technical Specifications

Transducer type	Condenser (back electret)
Operating principle	Pressure
Frequency response	20 - 20,000 Hz (50 - 16,000 Hz ±1.5 dB)
Polar pattern	Omnidirectional, diffuse field calibrated
Open circuit voltage at 1 kHz	15 mV/Pa (=-36.5 dBV) ±1 dB
Nominal impedance	160 Ω
Nominal load impedance	≥ 2.2 kΩ
Max. SPL at f = 1 kHz, k = 1%, R_L = 2.2 kΩ	122 dB_{SPL}
S/N ratio rel. to 1 Pa	>57 dB
A-weighted equivalent SPL	approx. 26 dB(A)
Power supply	12 - 48 V phantom supply
Current consumption	approx. 1.9 mA
Output	electronically balanced
Connection	3-pin XLR male
Dimensions:	
Length	133 mm
Shaft diameter	19/9 mm
Head diameter	9 mm
Weight (w/out cable)	88 g

Optional Accessories

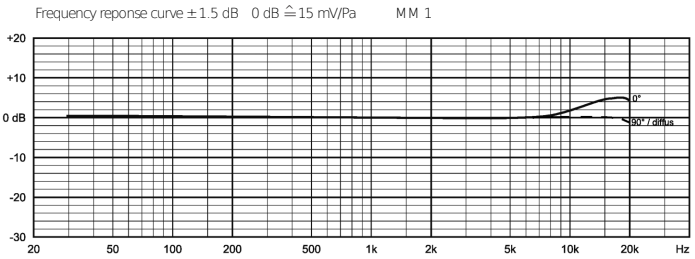
GST 400	Microphone stand, 3/8", height 0.90 - 1.65 m, with G 400 boom.	Order #421.294
GST 500	Microphone stand, 3/8", height 0.85 - 1.60 m, with telescopic G 500 boom.	Order #406.252
ST 400	Microphone stand, 3/8", height 0.90 - 1.65 mm.	Order #421.286
ST 500	Microphone stand, 3/8", height 0.85 - 1.60 mm.	Order #406.643
WS 10	Wind shield, charcoal grey	Order #403.008

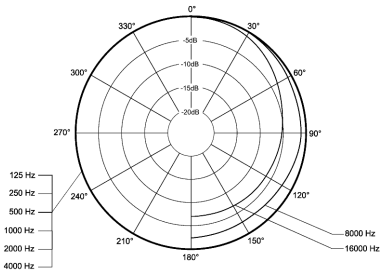
beyerdynamic GmbH & Co. KG
Theresienstr. 8 | 74072 Heilbronn – Germany
Tel. +49 (0) 71 31 / 617 - 0 | Fax +49 (0) 71 31 / 617 - 204
info@beyerdynamic.de | www.beyerdynamic.com

MM 1

Frequency Response & Polar Pattern

This polar pattern and frequency response curve (measuring tolerance ±1.5 dB) correspond to a typical production sample for this microphone. Each microphone is supplied with an individual 0° frequency response curve. Measured data can be requested via the following link: www.beyerdynamic.com/mm1-register





beyerdynamic GmbH & Co. KG
Theresienstr. 8 | 74072 Heilbronn – Germany
Tel. +49 (0) 71 31 / 617 - 0 | Fax +49 (0) 71 31 / 617 - 204
info@beyerdynamic.de | www.beyerdynamic.com

B Focusrite Octopre LE

SPECIFICATIONS

Mic Input Response

Gain = +13dB to +60dB
Input Impedance = 2.5k / 150 on Lo Z (Ch1 & 2)
EIN = 124dB @ 60dB Gain with 150 Termination & 22Hz/22kHz Filter
THD+N @ Min Gain (+13dB) = 0.0006% with 0dBu input & 22Hz/22kHz Filter
THD+N @ Max Gain (+60dB) = 0.003% with -36dBu input & 22Hz/22kHz Filter
THD+N @ Max Input (+9dBu) = 0.0008% with 22Hz/22kHz Filter
Frequency Response:
Min Gain (+13dB) with -13dBu input = -0.4dB @ 10Hz & -3dB @ 122kHz
Max Gain (+60dB) with -60dBu input = -2.3dB @ 10Hz & -3dB @ 67kHz
CMRR @ Max Gain (+60dB) = 80dB

Line Input Response

Gain = -10dB to +36dB
Input Impedance = 24k
Noise @ Unity Gain (0dB) = -88dBu with 22Hz/22kHz Filter
S/N Ratio relative to max headroom (+36dBu) = 124dB
S/N Ratio relative to 0dBfs (+22dBu) = 110dB
THD+N @ Unity Gain (0dB) = 0.001% with 0dBFS (+22dBu) input and 22Hz/22kHz Filter
Frequency Response @ Unity Gain (0dB) = -0.5dB @ 10Hz & -3dB @ 110kHz

Instrument Input Response

Gain = +13dB to +60dB
Input Impedance = 1M
Noise @ Min Gain (+13dB) = -87dBu with 22Hz/22kHz Filter
Noise @ Max Gain (+60dB) = -42dBu with 22Hz/22kHz Filter
THD+N @ Min Gain (+13dB) = 0.001% with 0dBu input & 22Hz/22kHz Filter
Frequency Response @ Min Gain (+13dB) with -13dBu input = -0.4dB @ 10Hz & -3dB @ 122kHz

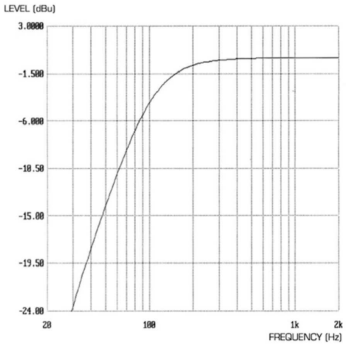
Input Meter

Peak Level Moving Coil Meter
-24dBfs to +2dBfs (-2dBu to +24dBu), +22dBu = 0dBfs
Overload LED's triggered @ 0dBfs (+22dBu)

High Pass Filter

Roll Off = 12dB/Octave 2 pole filter
Cut Off Frequency: -3dB @ 120Hz, -6dB @ 85Hz, -12dB @ 56Hz

Frequency Range:



OPTIONAL DIGITAL CONVERTER

DAC Performance

Playback Sample Frequency = 44.1kHz & 48kHz
Maximum Bit Depth = 24-bit
Maximum analogue output level = +22dBu
Dynamic Range = 107dB 'A' weighted

ADC Performance

Sample Frequency = 44.1kHz & 48kHz
Bit Depth = 24-bit
Maximum analogue input level = +22dBu (0dBfs)
Dynamic Range = 109.5dB 'A' weighted

Connections

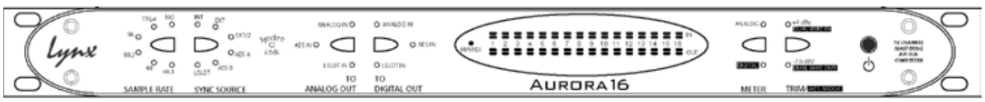
Digital in/out: ADAT -type optical 'lightpipe'
Wordclock in/out: BNC

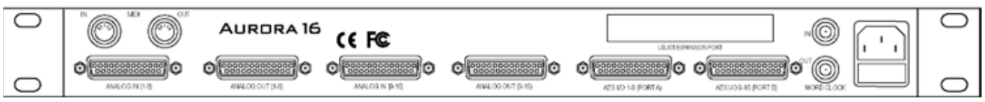
C Lynx Aurora 16 ADC/DAC

Lynx Aurora 16 and Aurora 8 Specifications

Analog I/O	
Aurora 8	Eight inputs and eight outputs
Aurora 16	Sixteen inputs and sixteen outputs
Type	Electronically balanced or unbalanced,
Level	+4 dBu nominal / +20 dBu max. or -10 dBV nominal / +6 dBV max
Input Impedance	Balanced mode: 24k Ω Unbalanced mode: 12k Ω
Output Impedance	Balanced mode: 100 Ω Unbalanced mode: 50 Ω
Output Drive	600 Ω impedance, 0.2 µF capacitance
A/D and D/A Type	24-bit multi-level, delta-sigma
Analog In Performance	
Frequency Response	20 Hz - 20 kHz, +0/-0.1 dB
Dynamic Range	117 dB, A-weighted
Channel Crosstalk	-120 dB maximum, 1 kHz signal, -1 dBFS
THD + N	-108 dB (0.0004%) @ -1 DBFS -104 dB (0.0006%) @ -6 DBFS 1 kHz signal, 22 Hz - 22 kHz BW
Analog Out Performance	
Frequency Response	20 Hz - 20 kHz, +0/-0.1 dB
Dynamic Range	117 dB, A-weighted
Channel Crosstalk	-120 dB max., 1 kHz signal, -1 dBFS
THD + N	-107 dB (0.00045%) @ -1 DBFS -106 dB (0.00050%) @ -6 DBFS 1 kHz signal, 22 Hz - 22 kHz BW
Digital I/O	
Number / Type	Aurora 8: 8 inputs and 8 outputs Aurora 16: 16 inputs and 16 outputs 24 bit AES/EBU format, transformer coupled
Channels	Aurora 8: 8 in/out in single-wire mode 4 in/out in dual-wire mode Aurora 16: 16 in/out in single-wire mode 8 in/out in dual-wire mode
Sample Rates	All standard rates and variable rates up to 192 kHz in both single-wire and dual-wire modes

On-board Digital Mixer (via AES16)	
Type	Hardware-based, low latency
Routing	Ability to route any input to any or multiple outputs
Mixing	Up to four input or playback signals mixed to any output, 40-bit precision
Status	Peak levels to -114 dB on all inputs and outputs
Connections	
Digital I/O Ports	25-pin female D-sub connectors Port A: channels 1-8 I/O Port B: channels 9-16 I/O (Aurora 16 only) Yamaha pinout standard
Analog I/O Ports	25-pin female D-sub connectors. Analog In 1-8 Analog In 9-16 (Aurora 16 only) Analog Out 1-8 Analog Out 9-16 (Aurora 16 only) Tascam pinout standard
External Clock	75 Ω BNC word clock input and output
MIDI	One input and one output. Standard opto-isolated, 5-pin female DIN connectors
Remote Control Options	
Function	Controls all I/O, levels, monitoring, routing and setting recall
Method	AES16/AES16*e*: With PC or Macintosh MIDI: Selected MIDI devices
General	
AC Power	100 / 115 / 230 VAC, 70 watts
Size	1.75" H x 19" W x 9" D
Shipping Weight	12 pounds
Certifications	CE and FCC Class B EMI, CE Product Safety
LSlot™ Expansion Port	
Compatibility	Supports Lynx LSlot expansion cards
Channels	Up to 16 input and 16 output simultaneously at up to 192 kHz sample rate
Optional Interface Cards for LSlot	
LT-ADAT	Provides 16-channel at 48 kHz, 8-channel at 96 kHz, 4-channel at 192 kHz ADAT Optical I/O
LT-HD	Provides interface for Avid® ProTools | HD® systems
LT-MADI	Provides up to 64 channels of I/O
LT-USB	Provides up to 16 channels of I/O, USB 2.0
LT-TB	Provides up to 32 channels of I/O





Lynx Studio Technology

Phone: 714-545-4700 Fax: 714-545-4777
Email: sales@lynxstudio.com
Website: http://www.lynxstudio.com

Aurora_11_14

Bibliography

[1] H. Teutsch, S. Spors, W. Herbordt, W. Kellermann, and R. Rabenstein, "An Integrated Real-time System for Immersive Audio Applications," in *IEEE Work. Appl. Signal Process. to Audio Acoust.*, no. 1, (New Paltz, NY, USA), pp. 1–4, IEEE, 2003.

[2] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Multiview Soundfield Imaging in the Projective Ray Space," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 6, pp. 1054 – 1067, 2015.

[3] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Soundfield imaging in the ray space," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 21, no. 12, pp. 2493–2505, 2013.

[4] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "The ray space transform: A new framework for wave field processing," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5696–5706, 2016.

[5] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[6] J. Ahrens, *Analytical Methods of Sound Field Synthesis*. Berlin, Germany: Springer, 1st ed., 2012.

[7] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London, UK: Academic Press, 1999.

[8] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Phd dissertation, Université Paris 6, 2001.

[9] A. J. Berkhout, D. De Vries, and P. Vogel, "Acoustic Control by Wave Field Synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764 – 2778, 1993.

[10] J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in *AES 114th Conv.*, (Amsterdam, NL), Audio Engineering Society, 2003.

[11] M. Neukom, "Ambisonic Panning," in *AES 123rd Conv.*, (New York, NY, USA), pp. 1–7, Audio Engineering Society, 2007.

[12] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. New York, NY, USA: Wiley & Sons, Inc., 1991.

[13] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice Hall, 1st ed., 2009.

[14] J. Lardies, H. Ma, and M. Berthillier, "Power estimation of acoustical sources by an array of microphones," in *Acoust. 2012 Nantes*, (Nantes, FR), pp. 2919–2924, HAL, 2012.

[15] T. Ajdler, *The plenacoustic function and its applications*. Phd, École Polytechnique Fédérale de Lausanne, 2006.

[16] "AES Standards on Acoustics - Sound Source Modeling - Loudspeaker Polar Radiation Measurements," Tech. Rep. Reaffirmed, New York, NY, USA, 2008.

[17] S. Qian and D. Chen, "Discrete Gabor Transform," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2429–2438, 1993.

[18] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part I)," *IEEE Signal Process. Mag.*, vol. 86, 2007.

[19] F. W. J. Olver, D. W. Lozier, F. Boisvert, Ronald, and C. W. Clark, *NIST Handbook of Mathematical Functions*. New York, NY, USA: National Institute of Standards and Technology, 2010.

[20] F. E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1," *J. Audio Eng. Soc.*, vol. 34, no. 4, pp. 227–235, 1986.

[21] F. M. Fazi, V. Brunel, P. A. Nelson, L. Hörchens, and J. Seo, "Measurement and Fourier-Bessel Analysis of Loudspeakers Radiation Patterns Using a Spherical Array of Microphones.," in *AES 124th Conv.*, (Amsterdam, NL), Audio Engineering Society, 2008.

[22] D. Queen, "The Effect of Loudspeaker Radiation Patterns on Stereo Imaging and Clarity," *J. Audio Eng. Soc.*, vol. 27, no. 5, pp. 368–379, 1979.

[23] E. Corteel and T. Caulkins, "Sound Scene Creation and Manipulation using Wave Field Synthesis," *Development*, p. 19, 2004.

[24] J. Ahrens and S. Spors, "Rendering of virtual sound sources with arbitrary directivity in higher order Ambisonics," in *AES 123rd Conv.*, vol. 1, (New York, NY, USA), pp. 65–73, Audio Engineering Society, 2007.

[25] J. Ahrens and S. Spors, "Implementation of Directional Sources in Wave Field Synthesis," in *IEEE Work. Appl. Signal Process. to Audio Acoust.*, no. 1, (New Paltz, NY, USA), pp. 1–4, IEEE, 2007.

[26] E. Verheijen, *Sound Reproduction by Wave Field Synthesis*. Phd, TU Delft, 1997.

[27] A. Kuntz and R. Rabenstein, "Cardioid Pattern Optimization for a Virtual Circular Microphone Array," in *EAA Symp. Auralization*, no. June, (Espoo, FI), pp. 1–4, EAA, 2009.

[28] A. Canclini, L. Mucci, F. Antonacci, A. Sarti, and S. Tubaro, "A methodology for estimating the radiation pattern of a violin during the performance," in *Eur. Signal Process. Conf.*, (Nice, FR), pp. 1546–1550, IEEE, 2015.

[29] K. Kowalczyk, O. Thiergart, M. Taseska, G. D. Galdo, V. Pulkki, L. Cristoferetti, and E. A. P. Habets, "Parametric Spatial Sound Processing," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 31–42, 2015.

[30] J. D. Johnston, J.-M. Jot, Z. Fejzo, and S. R. Hastings, "Beyond Coding: Reproduction of Direct and Diffuse Sounds in Multiple Environments," in *AES 129th Conv.*, (San Francisco, CA, USA), Audio Engineering Society, 2010.

[31] Michael M. Goodwin and Jean-Marc Jot, "Spatial Audio Scene Coding," in *AES 123th Conv.*, (San Francisco, CA, USA), Audio Engineering Society, 2008.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2nd ed., 2017.

[33] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 1st ed., 2010.

[34] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, 2015.

[35] M. Tohyama, A. Suzuki, and K. Sugiyama, "Active Power Minimization of a Sound Source in a Reverberant Space," in *IEEE Int. Conf. Acoust. Speech, Signal Process.*, (Glasgow, UK), pp. 2037–2040, IEEE, 1989.

[36] D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Active Acoustic Echo Cancellation in Spatial Soundfield Reproduction," in *IEEE Int. Conf. Acoust. Speech, Signal Process.*, (Vancouver, BC, CA), pp. 620–624, IEEE, 2013.

[37] Y. Suzuki, D. Brungart, H. Kato, K. Iida, and Y. Suzuku, *Principles and Applications of Spatial Hearing*. Singapore, SG: World Scientific, 1st ed., 2011.

[38] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Resolution issues in soundfield imaging: A multiresolution approach to multiple source localization," in *IEEE Work. Appl. Signal Process. to Audio Acoust.*, (New Paltz, NY, USA), pp. 1 – 5, IEEE, 2015.

[39] F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Extraction of acoustic sources for multiple arrays based on the Ray Space Transform," in *Hands-free Speech Commun. Microphone Arrays (HSCMA), 2017*, (San Francisco, CA, USA), pp. 146 – 150, IEEE, 2017.