Politecnico di Milano
Dipartimento di Elettronica, Informazione e Bioingegneria

Doctoral Program in Information Technology

A Unified Framework for Acoustic Scene Analysis, Synthesis, and Processing

Doctoral Dissertation of:
Lucio Bianchi
Candidate Id. 802176

Supervisor:
Prof. Augusto Sarti

Assistant Supervisor:
Prof. Fabio Antonacci

Tutor:
Prof. Andrea Virgilio Monti Guarnieri

The Chair of the Doctoral Program:
Prof. Andrea Bonarini

2015 - XXVIII cycle

A Unified Framework for Acoustic Scene Analysis, Synthesis and Processing
PhD thesis.

Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria

This thesis has been typeset by LaTeX.

Author's website: `home.deib.polimi.it/lbianchi`
Author's e-mail: `lucio.bianchi@polimi.it`

Abstract

This thesis proposes a unified framework that acts as a fundamental structure for a wide range of acoustic signal analysis, synthesis, and processing tasks. The framework emerges as a novel system of concepts and methods that encompasses the most recent advances in signal processing theory for a twofold purpose. From one point of view it provides the foundation for the design and the evaluation of new signal processing techniques based on conventional acoustic field representations; from the other point of view, it provides the required level of abstraction to introduce a novel representation of acoustic fields, based on the idea of decomposing them into a set of spatially confined wave objects. This thesis pursues both this purposes.

Conventional representations of acoustic field (in terms of, e.g., plane waves, spherical waves, etc.) provide a solid foundation for a wide range of acoustic signal processing tasks. Based on them, many techniques have been presented over the decades to solve a variety of problems. In this thesis, we interpret the representation of acoustic fields in terms of plane waves as an instance of multi-dimensional Fourier analysis. This interpretation is at the core of many techniques proposed in this thesis, as it supports the adoption of state-of-the-art signal processing techniques for the purpose of acoustic field analysis and synthesis. Moreover, the plane-wave representation leverages on a strong geometrical intuition: in this model, sound propagation is considered to occur along straight lines, referred to as *acoustic rays*. This fact enables the immediate understanding of several mechanism of acoustic propagation just employing basic geometric reasoning.

However, the plane-wave representation of acoustic fields is known to be, to some extent, limited: for instance, acoustic fields generated by near field sources are not conveniently represented by the superposition of plane waves. Therefore, a more general representation, allowing us to overcome the limitations of ray-based models, is in order. In this thesis, we define new elementary functions, referred to as *beams*, which are conceived as bundles of acoustic rays. Contrary to plane waves, which act as *global* descriptors for acoustic fields (since their spatial extension is infinite), beams are spatially confined; hence, the provide a *local* description of acoustic quantities.

Fourier theory shows its limits when trying to describe local phenomena. To overcome this limitation, in this thesis we embrace the mathematical theory of Gabor frames, which allows us to introduce a new representation of acoustic fields using beams as elementary constituents; this representation is referred to as *plenacoustic representation* and it parametrizes the acoustic field in terms of both spatial location and direction of propagation. In this setting, we derive a full-fledged *plenacoustic transform* that embeds the concept of a signal decomposition in the plenacoustic representation.

Thereafter, it emerges a promising global framework, in which acoustic fields are represented in a rigorous but extremely flexible and intuitive fashion. In evidence of this fact, in this thesis we propose techniques that, in our vision, will advance the state of the art of acoustic signal processing.

This page intentionally left blank.

Sommario

In questa tesi viene proposto un quadro concettuale che consente di unificare una varietà di operazioni di analisi, sintesi ed elaborazione di segnali acustici. Questo quadro concettuale è un sistema innovativo di concetti e metodi, e fa proprie le più recenti acquisizioni nell'ambito della teoria dell'elaborazione dei segnali con un duplice obiettivo: da un lato, infatti, fornisce i fondamenti per il progetto di nuove tecniche di elaborazione dei segnali basate su rappresentazioni convenzionali dei campi acustici; dall'altro, fornisce un livello di astrazione adeguato per introdurre nuove rappresentazioni di campi acustici, basate sull'idea di scomporre i campi in un insieme di oggetti d'onda confinati spazialmente.

Partendo dalle rappresentazioni convenzionali dei campi acustici (in termini, ad esempio, di onde piane, onde sferiche, etc.), che forniscono un solido fondamento per molte recenti tecniche di elaborazione dei segnali acustici, in questa tesi si interpreta la rappresentazione dei campi acustici in termini di onde piane come un'istanza dell'analisi di Fourier multidimensionale. Questa interpretazione è centrale in molte tecniche proposte in questa tesi, dal momento che supporta l'adozione delle più innovative ed efficienti tecniche di elaborazione dei segnali con lo scopo di analizzare e sintetizzare campi acustici. Inoltre, la rappresentazione in onde piane fa leva su una forte intuizione geometrica: in questo modello, la propagazione del suono è descritta tramite rette, dette *raggi acustici*. Questo consente la comprensione immediata di molti meccanismi di propagazione acustica, utilizzando solamente ragionamenti geometrici basilari.

Tuttavia, la rappresentazione di campi acustici in onde piane è nota per essere, in una certa misura, limitata: ad esempio, i campi acustici generati da sorgenti in campo vicino non sono rappresentati in modo conveniente da una sovrapposizione di onde piane. Per questo motivo, nella tesi viene costruita una rappresentazione più generale, che consente di superare le limitazioni dei modelli basati sui raggi acustici. A questo scopo vengono definite nuove funzioni elementari, dette *beam*, che sono concepiti come fasci di raggi acustici. Contrariamente alle onde piane, che sono descrittori *globali* per i campi acustici (dal momento che la loro estensione spaziale è infinita), i beam sono spazialmente confinati, quindi forniscono una descrizione *locale* delle quantità acustiche.

La teoria di Fourier mostra, però, i propri limiti nel cercare di descrivere fenomeni locali. Per superare questa limitazione, in questa tesi si abbraccia la teoria matematica dei *frame* di Gabor, la quale consente di introdurre una nuova rappresentazione per i campi acustici, utilizzando i beam come elementi costitutivi. Questa rappresentazione è detta *rappresentazione plenacustica* e parametrizza il campo acustico in termini sia della posizione spaziale che della direzione di propagazione. In questo contesto, si deriva la *trasformata plenacustica* che realizza il concetto della scomposizione dei segnali acustici nella rappresentazione plenacustica.

Da questo emerge un promettente quadro concettuale, in cui le quantità acustiche sono rappresentate in modo rigoroso ma estremamente flessibile e intuitivo. Grazie a ciò, in questa tesi si propongono tecniche che, nelle intenzioni degli autori, avanzeranno lo stato dell'arte dell'elaborazione dei segnali acustici.

This page intentionally left blank.

Acknowledgments

This thesis is the result of a three-year collaboration with the Image and Sound Processing Group (ISPG) at Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, lead by Prof. Augusto Sarti and Prof. Stefano Tubaro. It has been a real honor for me to work in such an inspiring and collaborative research environment.

I would like to thank my supervisor, Prof. Augusto Sarti, who has been a great source of inspiration and motivation from my first day on; he has been greatly supportive and helpful at every stage of my work: I am very grateful to Augusto for all his help. A special thank to Prof. Fabio Antonacci, who trained me, with infinite patience, in the art of being pragmatic about research, focusing on realistic goals and make things work; I would like to thank Fabio for all its invaluable help during my work. My gratitude also goes to Dr. Antonio Canclini and Dr. Dejan Marković, whose studies are the solid roots of my work. I would like to thank all my friends at ISPG for all the good time spent together: a big thank goes to Antonio, Bruno, Daniele, Michele, Dejan, Fabio, Francesco, Massimiliano, Muhammad, Silvia, and Paolo.

I'm grateful to Dr. Filippo Maria Fazi for welcoming me into the Virtual Acoustics and Audio Engineering group at the Institute of Sound and Vibration Research, University of Southampton; I would like to thank Filippo for all his help, his hospitality and for the wonderful discussions on acoustic theory.

My gratitude goes to Prof. Francesco Piazza and Prof. Rudolf Rabenstein for accepting to review this thesis, for their constructive comments and helpful suggestions.

Thanks to my parents Franco and Giovanna, who gave me all the opportunities one can wish for. Finally, a special dedication goes to Fosca, who, in her effort to keep my mind to better things than work, makes life beautiful.

Milano, 2015 L. B.

This page intentionally left blank.

Contents

This page intentionally left blank.

List of Figures

Acronyms

DAS Delay and Sum

DI Directivity Index

DNR Diffuse-to-Noise Ratio

DoA Direction of Arrival

HOA Higher-Order Ambisonics

MEMS Micro Electro-Mechanical Systems

MVDR Minimum Variance Distortionless Response

PSF Point Spread Function

RLS Robust Least Squares

SDM Spectral Division Method

SIR Signal-to-Interference Ratio

SNR Signal-to-Noise Ratio

SRP Steered Response Power

TDoA Time Difference of Arrival

ToA Time of Arrival

WFS Wave Field Synthesis

WNG White Noise Gain

This page intentionally left blank.

Nomenclature

Sets, vectors spaces and Hilbert spaces

$\mathbb{C}$	The set of complex numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integer numbers
$\mathbb{C}^{\mathbb{R}}$	Vector space of complex-valued functions of a single real variable
$\mathbb{C}^{\mathbb{R}^D}$	Vector space of complex-valued functions of D real variables
$\mathbb{C}^{\mathbb{Z}}$	Vector space of complex-valued sequences of a single integer variable
$\mathcal{L}^2(\mathbb{R})$	Hilbert space of finite energy function
$\ell^2(\mathbb{Z})$	Hilbert space of finite energy sequences

Constants

c	Speed of sound
j	Imaginary unit (i.e. $j = \sqrt{-1}$)

Operations and operators

$*$	Convolution
$\circledast$	Circular convolution
$(\cdot)^*$	Complex conjugation
$(\cdot)^H$	Conjugate transpose
$(\cdot)^T$	Transposition
$\langle \cdot, \cdot \rangle$	Inner product
$\lvert \cdot \rvert$	Absolute value
$\lVert \cdot \rVert$	$\mathcal{L}_2$-norm
$\lVert \cdot \rVert_F$	Frobenius norm
$E[\cdot]$	Mathematical expectation
$\mathrm{diag}(\cdot)$	Operates on a vector and forms a diagonal matrix

$\mathrm{tr}(\cdot)$ Sum of the diagonal elements of a square matrix

∇^2 Laplace operator

∇_T^2 Transverse Laplace operator, see equation (3.99)

Special functions and sequences

$\delta(\cdot)$ Dirac's delta function [Vetterli et al., 2014, pp. 316-317]

$\delta[k]$ Kronecker's sequence [Vetterli et al., 2014, p. 186]

$G(\mathbf{r}|\mathbf{r}', \omega)$ Green's function from a source in $\mathbf{r}'$ to an observer in $\mathbf{r}$ at frequency ω

$H_m^{(1,2)}(\cdot)$ Hankel function of first and second kind [Olver, 2010, p. 217]

$h_l^{(1,2)}(\cdot)$ Spherical Hankel function of first and second kind [Olver, 2010, p. 262]

$J_m(\cdot), Y_m(\cdot)$ Bessel's function of first and second kind [Olver, 2010, p. 217]

$j_l(\cdot), y_l(\cdot)$ Spherical Bessel function of first and second kind [Olver, 2010, p. 262]

$\mathsf{P}_l^m(\cdot)$ Associated Legendre Polynomial of degree l and order m [Olver, 2010, p. 360]

$\mathsf{P}_\mu(\cdot)$ Legendre polynomial of degree μ [Olver, 2010, p. 440]

$Y_l^m(\theta, \phi)$ Spherical harmonic of degree l and order m [Ahrens, 2012, p. 27]

Variables

$\mathbf{k} = [k_x, k_y, k_z]$ Wavenumber vector

λ Wavelength

$\hat{\mathbf{n}}(\mathbf{r})$ Normal unit vector at $\mathbf{r}$

ϕ Azimuth angle

$p(\mathbf{r}, t)$ Acoustic field at point $\mathbf{r}$ and time t

$P(\mathbf{r}, \omega)$ Acoustic field at position $\mathbf{r}$ and temporal frequency ω

$\mathbf{r} = [x, y, z]^T$ Vector of spatial coordinates

$\hat{\mathbf{r}}$ Unit vector in the direction of $\mathbf{r}$

t Time [s]

θ Co-elevation angle

ω Angular frequency [$\mathrm{rad\,s^{-1}}$]

Chapter 1

Introduction

The ability of sensing spatial properties of sound enables humans to experience immersivity: sound provides information that enforces the sense of presence in a scene. As a matter of fact, sound carries semantic information both on the geometric structure of the surrounding environment and on the position of the listener inside the scene. If we remove all spatial attributes to sound, the listener experiences a dramatic impairment, and the sense of presence is inhibited.

Only in the past two decades has the signal processing community started to consider sound signals as functions of both space and time. In the following of this thesis, we refer to the class of sound signals that are functions of space and time as *acoustic signals*. The power of this idea, foreseen by Rabenstein et al. [2005], Ajdler [2006], Ajdler et al. [2006a], Rabenstein et al. [2006], Kuntz [2008], Annibale et al. [2009], lies in the fact that it enables the use of well-established and computationally efficients methods from multi-dimensional signals and systems theory [Rabenstein and Trautmann, 2002, 2003] when dealing with acoustic signals, explicitly exploiting the information carried by the spatial structure of the observed signals.

Spatial distributions of microphones and loudspeakers provide the technological mean for capturing and reproducing acoustic signals. In the following, the term *array* is used with reference to a regular arrangement of acoustic transducers. Just for the moment, we focus our attention to microphone arrays, since their interpretation is easier; however, it is important to notice that the very same reasoning is applied in this thesis also to loudspeaker arrays. Microphone arrays, as they are conventionally employed, are conceived as *global* sources of information, in the sense that their spatial extension is exploited to extract information about the acoustic field at a single point (the center of the array, or its reference point, conventionally). Devices such as acoustic cameras [Brooks and W. M. Humphreys, 2004, Yardibi et al., 2008, O'Donovan et al., 2008, Legg and Bradley, 2013] and spherical microphone arrays [Meyer and Elko, 2002, Rafaely, 2005, Jin et al., 2014] emerged from this view: they do not attempt to extract punctual information at microphone capsules, rather they rely on their spatial extension to extract information that is valid only at its center. If the acoustic field can be considered "uniform" in the region occupied by the array – or equivalently, if the array can be considered compact –, many effective techniques have been proposed in the literature to extract salient information from array data.

However, there are situations in which the acoustic field evolves in space: this is the typical case of sound sources placed in the near field of the microphone array; in such a case, the spatial properties of the acoustic field vary if one moves along the array. To tackle this scenario, it is useful to adopt the concept of the *plenacoustic function*, which has

been introduced in the literature by Kubovy and Valkenburg [2001], Ajdler and Vetterli [2003a,b], Ajdler et al. [2005, 2006b,a], Ajdler [2006], for the purpose of encoding the spatial evolution of the acoustic field as a function of the position along the array. More recently, the concept of the plenacoustic function has been revisited by Marković et al. [2012, 2013a, 2015] in a *directional* form, as describing the acoustic field parametrized by spatial location and direction of propagation; this form of the plenacoustic function is referred to as the *directional plenacoustic function*. The directional plenacoustic function is the representation of election to encode *local* information on the acoustic field, as it allows us to rethink the use of arrays in a local fashion, i.e. by subdividing the array into smaller portions and by processing their data from a local perspective. This is the route followed in this thesis.

In this thesis we extend the globality and locality concepts also to arrays of loudspeakers, thanks to the reciprocity principle.

1.1 Background

The discipline that deals with the processing and extraction of information from acoustic signals is named *acoustic signal processing*. This area of signal processing is currently seeing an increased level of interest in both the academy and the industry. It encompasses the fundamental theory, applications and algorithms suitable to deal with acoustic signals.

Parametric Representations. Acoustic signal representations are at the heart of most acoustic signal processing techniques. Two representation paradigms have emerged in the recent literature. One is based on a *parametric* description of the acoustic field [Faller, 2004, Pulkki, 2007, Goodwin and Jot, 2008, Ahonen et al., 2008, Laitinen and Pulkki, 2009, Schultz-Amling et al., 2010, Herre et al., 2011, Thiergart et al., 2013, 2014, Primavera et al., 2014, Kowalczyk et al., 2015]; within this paradigm, audio signals corresponding to each sound event are considered along with, possibly, many parameters that describe the acoustic scene itself (e.g. source locations, environment geometry, etc., depending on the amount of prior information available and on the desired accuracy).

Microphone arrays have traditionally been exploited to identify the Direction of Arrival (DoA) of a sound source and enhance sounds coming from desired directions [Mitlanoudis and Davies, 2003, Markovich et al., 2009, Hioka and Betlehem, 2013, Thiemann and Vincent, 2013]. This task is accomplished by appropriately filtering microphone signals, thus achieving directional sound reception [Kennedy et al., 1998, Ryan, 1998, Ward et al., 1998, Fiori and Piazza, 2000, Meyer, 2001, Benesty et al., 2007, Guillaume and Grenier, 2007, Li and Duraiswami, 2007, Atkins, 2011, Gauthier et al., 2011, Li and Chen, 2011, Lorente et al., 2011, Pirro et al., 2011, Yan et al., 2011, Spors et al., 2012, Pirro et al., 2012]. This filtering operation is known as *beamforming* [Veen and Buckley, 1988], a concept borrowed from the literature on antenna arrays, and is based on the idea of altering microphone signals in such a way that sounds coming from desired angles experience constructive interference (thus being enhanced), while others experience destructive interference (thus being attenuated).

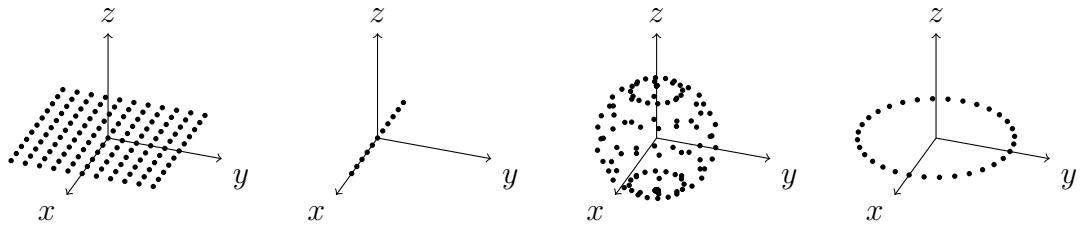
Another important application is the localization of a sound source, i.e. the estimation of its position given measurements of the acoustic field [Schmidt, 1972, Stoica and Li, 2006]. Many methods have been proposed for that purpose, mainly based on the estimation of Time Difference of Arrival (TDoA), e.g. methods in Woodward [1964], Cook and Bernfeld [1967], Altes [1979], Adams et al. [1980], Carter [1981], Chan and Ho [1994], Smith and Abel [1987], Strobel and Rabenstein [1999], Strobel et al. [1999], Strobel and

Rabenstein [2000], Strobel et al. [2001], Annibale [2010], Annibale and Rabenstein [2010, 2012], Compagnoni et al. [2012], Bestagini et al. [2013], Canclini et al. [2013], Compagnoni et al. [2014], Canclini et al. [2015]; another class of methods is based on the estimation of the Steered Response Power (SRP), as in Omologo and Svaizer [1994], Gillette and Silverman [2008], Wei and Ye [2008].

Another application field that is gaining increasing interest is the inference of environment geometry using both single measurements [Marković et al., 2013b, Zamaninezhad et al., 2014] and microphone arrays [Canclini et al., 2011, Antonacci et al., 2012, Dokmanić et al., 2013]: adopting suitable measurement techniques and then analyzing the temporal and spatial correlation between echoes, it has been shown that it is possible to estimate several geometric parameters of the environment, like dimension of the room, location of walls, reflectors, etc., in an acoustic-only fashion. Information on the environment can be exploited, for instance, for the equalization, correction and compensation of room acoustics, as in Omura et al. [1999], Spors et al. [2003], Spors [2005], Spors et al. [2007], Canclini et al. [2012], Cecchi et al. [2012], Carini et al. [2012], Cecchi et al. [2013a,b, 2014a,b,c], Canclini et al. [2014a,b], Poletti et al. [2015].

Nonparametric Representations The other paradigm is based on a *nonparametric* description, where no a-priori knowledge on the acoustic scene is assumed [Schetelig and Rabenstein, 1998, Buchner et al., 2002, Herbordt et al., 2003, Teutsch et al., 2003, Pinto, 2010, Pinto and Vetterli, 2010, Pinto et al., 2014]. Within this paradigm, the sound scene is described by the acoustic field itself, considered as a function of space and time. Of course, adopting a point-by-point description for the acoustic field would require to consider an amount of data that is impossible to handle in any practical application. For this reason, scholars developed acoustic signal representations that efficiently encode all the information contained in the acoustic field. Representative examples of these physically-motivated representations are the *plane-wave representations*, as introduced by Whittaker [1903], Weyl [1919], Lalor [1968], Sherman [1969, 1972], Devaney and Sherman [1973], Devaney and Wolf [1974], Nieto-Vesperinas [1988], Ergin et al. [1999], Hansen and Yaghjian [1999], Perrey-Debain [2006], Moiola et al. [2011], Hansen [2014] and adopted in the acoustic signal processing context, among the others, by Rafaely [2004], Duraiswami et al. [2005], Park and Rafaely [2005], Günel et al. [2007], Rafaely et al. [2007], Zotkin et al. [2010], Peled and Rafaely [2013]; the *spherical wave* representation [Devaney and Wolf, 1974, Mohlenkamp, 1999, Meyer, 2001, Abhayapala and Ward, 2002, Meyer and Elko, 2002, MacPhie and Wu, 2003, Gover, 2005, Poletti, 2005, Rafaely, 2005, Li and Duraiswami, 2007, McEwan et al., 2007, Rafaely, 2008, Rafaely and Kleider, 2008, Zotkin et al., 2008, Koretz and Rafaely, 2009, Parthy et al., 2009, Zotter, 2009, Costa et al., 2010, Fisher and Rafaely, 2011, Rafaely and Khaykin, 2011, Yan et al., 2011, Ahrens and Spors, 2012, Lai et al., 2013, Jin et al., 2014, Khalid et al., 2014, Zhang et al., 2015]; and the cylindrical wave representation [Cincotti et al., 1993, Rabenstein et al., 2006, Teutsch and Kellermann, 2006, Kuntz and Rabenstein, 2007, Ahrens and Spors, 2008a, Abhayapala and Gupta, 2010, Poletti, 2010, Kolundzija et al., 2011, Hioka and Betlehem, 2013, Alexandridis et al., 2013, Askari et al., 2013, Thomas et al., 2014, Hoffmann and Fazi, 2015].

Nonparametric analysis concerns the problem of estimating the coefficients of a set of basis functions in which the acoustic field can be decomposed. This is accomplished by sampling the acoustic field using an array of microphones, placed in space to capture variations of the acoustic field as a function of space. Of course, the naive idea of placing tightly spaced microphones to cover the whole spatial extension of the acoustic field is not feasible in any practical sense. Indeed, both researchers and industry devised specific



(a) Planar grid. **(b)** Linear. **(c)** Spherical. **(d)** Circular.

Figure 1.1: Most common microphone array configurations.

microphone configurations that enable to capture relevant spatial information carried by the acoustic field. Figure 1.1 shows some illustrative configurations for microphone arrays: a planar grid (Fig. 1.1a), a linear configuration (Fig. 1.1b), a spherical shell (Fig. 1.1c) and a circular deployment (Fig. 1.1d). Recent works (e.g. [Ajdler et al., 2006a]) have introduced the concept that signals from multiple microphones can be interpreted as time-space samples characterized by specific temporal and spatial sampling frequencies, the first being determined by the temporal interval between adjacent time samples, while the latter being determined by the spacing between microphones. In general, the spatial dependency of an acoustic field can be represented in terms of specific sets of basis functions (plane waves, spherical waves, cylindrical waves, etc.). Many acoustic signal processing techniques have been proposed to decompose an acoustic field in terms of these basis functions, e.g. the techniques in Rafaely [2005], Guillaume and Grenier [2007], Kuntz [2008], Yan et al. [2011], Gauthier et al. [2011], Samarasinghe et al. [2014].

Another important application for nonparametric representation is the synthesis of acoustic field, usually referred to as *rendering*. Acoustic rendering is conceived as the process of synthesizing physically accurate acoustic fields over an extended area [Spors et al., 2004], [Spors et al., 2013, p. 1920]. The rendering of acoustic fields is a goal that has been pursued by the acoustic signal processing community for decades. Solutions based on spatial arrangements of loudspeakers (*loudspeaker arrays*) constitute today a well-established technology. In broad terms, what such solutions do is to reproduce a desired acoustic field through the superposition of elementary contributions emitted by the loudspeakers. Many techniques have emerged in order to obtain suitable loudspeaker driving signals, ranging from analytical (e.g. in Berkhout et al. [1993], Poletti [2005], Wu and Abhayapala [2009], Ahrens and Spors [2010], Gupta and Abhayapala [2011]), to numerical approaches (e.g. in Kirkeby et al. [1996], Ward and Abhayapala [2001], Lilis et al. [2010]). Considering analytical approaches, the most representative acoustic rendering methods are known as Wave Field Synthesis (WFS) [Berkhout et al., 1993, Spors et al., 2008] and Higher-Order Ambisonics (HOA) [Daniel, 2003, Poletti, 2005, Ahrens and Spors, 2008b,a, Wu and Abhayapala, 2009]. Being based on the decomposition of the acoustic field into spherical waves, the original formulation of HOA is constrained to circular or spherical loudspeaker arrays. A successive extension, known as Spectral Division Method (SDM) is introduced by Ahrens and Spors [2010], and it exploits the same principles of HOA adopting linear or planar loudspeaker arrays.

Geometric Representations. Another class of representations emerges by moving from the realm of wave acoustics to the realm of geometric acoustics. Geometry-inspired methods rely on a simplified model of acoustic propagation, where acoustic information is transported along straight lines, the *acoustic rays*. Although, from a physical perspective,

this view provides a rough approximation of an acoustic field (especially in the near field of sound sources), the ray-based model proves its importance for the modeling of arbitrarily complex acoustic environments, for which a physically accurate modeling is not practical.

Recently, Antonacci et al. [2008] introduced the concept of the *ray space*, defined as a projective domain whose primitives are acoustic rays. This domain is built on a powerful idea, according to which points in the ray space represent plane-wave components of the acoustic field; moreover, the adoption of a projective domain opens the door to the use of fast techniques from computational geometry. The power of the ray space is demonstrated by Marković et al. [2013a, 2015], where it is employed as the domain of the directional plenacoustic function. In those works, authors proves that all acoustic objects appear in the ray space as linear patterns. These linear patterns are easy to deal with in a wide range of problems: localization of multiple sources [Marković et al., 2013a], radiometric analysis [Marković et al., 2014], environment inference [Antonacci et al., 2012], etc. All the problems that can be envisioned in the ray space are suitable to be solved with pattern processing algorithms. However, the plenacoustic representation in the ray space has not been embedded, so far, with the concept of a signal decomposition in it. The ultimate goal of this thesis is to define such concept, combining the computational ease and the intuition enabled by geometry-inspired representations with the accuracy of physically-motivated representations.

1.2 Goals and Methodology

Goals. We have already raised the issue of global versus local paradigms for array processing. In all situations in which the spatial properties of the acoustic field can be considered "uniform" over the whole extension of the array, the global processing paradigm offers a valuable tool to design effective and efficient signal processing solutions. The first goal that we pursue in this thesis is to enhance those solutions. We do that by adopting signal processing methodologies based on Fourier theory, which allows us to devise techniques that exhibit high accuracy while being conceived to work with real-world array data. This is the perspective adopted in Part II of this thesis.

In Part III of this thesis we intend to move to an higher degree of proposed innovation, shifting our mindset towards a local paradigm for array processing. In order to do that, we consider portions of a whole array: this step introduces resolution issues since, for each portion of the array, we rely on less data and on a reduced spatial extension. However, we capture far more information on the acoustic field, being able to represent variations of the spatial structure of the acoustic field itself. Therefore, we set a trade-off between the achievable resolution and the capability of capturing the local evolution of the acoustic field. We intend to provide some tools, based on the mathematical theory of redundant representation of signals, that allow us to find the optimal trade-off between resolution and locality.

In this thesis we aim at rethinking array processing in these terms. We derive a new representation of acoustic fields, which emerges from a generalized version of Fourier theory that, based on the mathematical theory of frames, allows us to define a spatially local Fourier transform. We show that this signal processing abstraction is supported by a strong physical intuition. As a matter of fact, we define acoustic beams as the elementary wave objects at the roots of the spatially local Fourier transform.

Methodology. In order to pursue these goals, in this thesis, we devise and we follow a three-step methodology.

1. First, we identify a suitable set of physical models, ranging from conventional wave objects (i.e. plane waves, spherical waves, and cylindrical waves) to more geometrically intuitive models (i.e. acoustic rays). Beams emerge as wave objects that enable a physically accurate representation of acoustic phenomena maintaining, at the same time, the strong geometric intuition behind acoustic rays.

2. On top of these physical models, we derive specific representations of acoustic fields: this is accomplished by merging tools from mathematical physics and from the theory of signal representations. According to the specific wave object assumed as the basis for the representation, we end up with representations that are, in general, infinite-dimensional and continuous.

3. By introducing some assumptions, we can reduce infinite-dimensional and continuous representations to finite-dimensional and discrete ones. In this latter setting, we can rely on the powerful tools provided by linear algebra in order to define signal processing operations (e.g. filtering, spectral analysis, etc.) in the context of such representations. Furthermore, the properties of linear algebra allow us to discover and derive fast algorithms for the implementation of those processing operations.

The methodology outlined above, in our view, paves the way to the establishment of the unified signal processing framework envisioned in this thesis.

1.3 Thesis Outline and Main Contributions

This section provides a chapter by chapter overview, summarizing the main contributions of this work. References to the publications that have been produced in the course of the work are provided.

Part I provides necessary background information for later chapters. In particular, Chap. 2 provides a review of signals and signal representations in the algebraic setting of Hilbert spaces. In this setting, *bases* are introduced as non-redundant representations of signals, and *frames* as their redundant counterpart. Fourier and Gabor representations are introduced as specific instances of signal representations.

Chapter 3 reviews the physical laws governing sound generation and propagation; the wave equation and its fundamental solutions (plane waves, spherical waves, cylindrical waves) are introduced, as they constitute the basic wave-oriented modeling tools used in acoustical signal processing. The chapter follows with a presentation of ray acoustics, derived from approximations to the wave equations. Finally, the paraxial approximation to the wave equation is thoughtfully discussed, as it provides the foundation to beam acoustics. The paradigm of beam acoustics is, at our knowledge, a new introduction in the acoustic signal processing literature and it provides the root for the convergence of wave acoustics and geometric methods envisioned in this thesis.

Tools from signal processing and acoustic theory are then merged in Chap. 4 to provide a necessary background on the acoustic field representations widely adopted in the literature. A unified review of representations based on plane waves, spherical waves, and cylindrical waves is provided. In this context, the foundations of beam acoustics introduced in Chap. 3 are employed for the purpose of introducing a novel full-fledged representation of acoustic signals.

In Part II, we present new techniques for the analysis and synthesis of acoustic fields based on conventional representations. In particular, Chap. 5 considers the problem of analyzing acoustic fields using a compact spherical array of microphones. With more

details, a spherical microphone array is employed to analyze the plane wave DoAs and strengths, based on high-resolution tools from the literature on spectral analysis of signals. This technique is validated against the practical problem of identifying acoustic reflections in arbitrarily complex environments. The work discussed in this chapter has been presented in Bianchi et al. [2015a].

Chapter 6 considers the problem of synthesizing a physically accurate acoustic field in an area surrounded by an array of loudspeakers. For this purpose, we present a technique based on the decomposition of the acoustic field in terms of propagating plane waves; the proposed technique can be readily applied to arbitrary distributions of loudspeakers. Moreover, we consider the problem of operating the loudspeaker system in real-world environments, many non-idealities impair the obtained results. In order to reduce the impact of non-idealities, a technique is presented that provides an higher accuracy even in the presence of errors in the modeling of the propagation. The work discussed in this chapter has been presented in Bianchi et al. [2014, 2016].

Part III presents the most innovative contributions of this thesis. Chap. 7 presents the directional plenacoustic representation in the light of local Fourier analysis. In this view, two techniques are presented: one addresses the problem of localizing multiple acoustic sources with high accuracy; the second relies on this information to enhance the signal emitted from a specific source, while attenuating other interferers and noise. Thereafter, Chap. 7 continues by providing the theoretical foundation for a full-fledged acoustic field representation in the domain of the directional plenacoustic function; the convergence between local Fourier analysis methods and the beam-based representation of acoustic field provides the root for such novel representation. The work discussed in this chapter has been partially presented in Bianchi et al. [2013, 2015b].

Finally, Chap. 8 draws some conclusions and gives directions for further research.

This page intentionally left blank.

Part I

Preliminaries

This page intetntionally left blank.

Chapter 2

Signals and Representations

This chapter presents a review of signals and signal representations. Signals are conceived here as functions defined on a proper domain (continuous or discrete, one-dimensional or multi-dimensional), which encode information on some relevant phenomena. In the context of this thesis, these phenomena are acoustic phenomena; however, the concepts and the methodologies reviewed in this chapter are valid for any variety of signals. The abstraction of the signal from its specific physical interpretation is a great advantage provided by the signal processing paradigm. Indeed, it allows to define concepts and methodologies independent on the physical nature of specific signal at hand. This constitutes a first step towards the unifying paradigm addressed in this thesis.

Following Vetterli et al. [2014], signals and their representations are introduced here through the machinery of Hilbert spaces. Hilbert spaces are abstract vector spaces that enable to extend in a straightforward fashion all the notions valid in the common Euclidean space to spaces with infinite dimensions, both countable and uncountable. Thus, Hilbert spaces provide a common conceptual framework to deal with wide classes of signals, being them continuous, discrete, one-dimensional, and multi-dimensional.

An Hilbert space is defined, in this thesis, as a vector space in which

- the operation of *inner product* is defined;
- the inner product induces a *norm*;
- the notion of *completeness* applies.

Only an operative perspective on the machinery enabled by Hilbert spaces will be provided in this thesis; the interested reader is referred to Luenberger [1969], Kennedy and Sadeghi [2013], Vetterli et al. [2014] for a complete overview of the application of Hilbert spaces in signal processing. The primary reference for the material reviewed in this chapter is Vetterli et al. [2014], that features a good balance between mathematical rigor and intuition.

Hilbert spaces provide an abstract mean to leverage on geometric reasoning when processing signals. Indeed, *signal representations* can be conceived as coordinate systems for the Hilbert space, and transitions among different signal representations appear as *projections*. *Basis* are introduced as complete and non-redundant representations, and *frames* as their complete and redundant counterparts. In this setting, an arbitrary signal can be represented as a linear combination of prototype signals.

In this chapter we introduce the signal spaces use din the rest of this thesis. We introduce signal spaces whose domain is either *continuous* or *discrete*. In each case, the appropriate definitions for the inner product, for the norm and for the fundamental

operation of *convolution* are reviewed. Moreover, we introduce Fourier transformations for each signal class as a projection onto a (non-redundant) set of specific prototype signals: complex exponentials.

2.1 Signals

This section introduces the signal classes that will be used in the rest of this thesis. First, one-dimensional signals, whose domain is a continuum, are introduced in both the aperiodic and periodic settings; then the same concepts and tools are introduced for discrete (aperiodic and periodic) signals. Finally, the extension to the multi-dimensional realm is provided. This section acts as an introduction for the notation adopted in the rest of this thesis.

When dealing with one-dimensional signals, although the setting of this chapter is general and abstract (in the sense that here the physical interpretation for the signals is not considered), the classical perspective that associates the independent variables $t \in \mathbb{R}$ or $l \in \mathbb{Z}$ with *time* is followed. This choice is motivated by the fact that the overall nomenclature pertaining one-dimensional signals has, historically, been developed with time-domain signals in mind (see, e.g., Oppenheim et al. [1998]). On the other hand, when considering multi-dimensional signals the variable $\mathbf{r} \in \mathbb{R}^D$ is associated to *space*[1]. The rigorous interpretation of time and space variables is postponed to Chap. 3.

2.1.1 One-Dimensional Continuous Signals

The set of complex functions of a single continuous variable, defined on the domain of real numbers, (along with function addition and multiplication by a scalar) forms the vector space $\mathbb{C}^{\mathbb{R}}$ [Vetterli et al., 2014, p. 19]. Signals in $\mathbb{C}^{\mathbb{R}}$ are denoted by

$$f(t), \quad f : \mathbb{R} \to \mathbb{C}. \tag{2.1}$$

Real-valued signals (i.e. signals in $\mathbb{R}^{\mathbb{R}}$) are considered as a subset of $\mathbb{C}^{\mathbb{R}}$. The standard inner product in $\mathbb{C}^{\mathbb{R}}$ between two functions $f_1(t)$ and $f_2(t)$ is [Vetterli et al., 2014, p. 24]

$$\langle f_1, f_2 \rangle = \int_{-\infty}^{\infty} f_1(t) f_2^*(t)\, dt, \tag{2.2}$$

where $(\cdot)^*$ denotes complex conjugation. Equation (2.2) induces the $\mathcal{L}^2$ (Euclidean) norm [Vetterli et al., 2014, p. 28]

$$\|f\| = \sqrt{\langle f, f \rangle} = \left(\int_{-\infty}^{\infty} |f(t)|^2\, dt \right)^{1/2}, \tag{2.3}$$

which is closely related to the *energy of a signal*, being the energy of the signal f defined as $\|f\|^2$. In the following, the attention is restricted to the space of *finite-energy* signals $\mathcal{L}^2(\mathbb{R})$, for which the $\mathcal{L}^2$ norm (2.3) is finite, i.e. $\|f\| < \infty$. The constraint of a finite energy turns the vectors space $\mathbb{C}^{\mathbb{R}}$ into the Hilbert space $\mathcal{L}^2(\mathbb{R})$ [Vetterli et al., 2014, pag. 346].

The convolution between two functions $f_1(t)$ and $f_2(t)$ is defined as in Vetterli et al. [2014, p. 356]

$$(f_1 * f_2)(t) = \int_{-\infty}^{\infty} f_1(\tau) f_2(t-\tau)\, d\tau = \int_{-\infty}^{\infty} f_1(t-\tau) f_2(\tau)\, d\tau. \tag{2.4}$$

[1] In this thesis vector quantities are denoted with a lowercase bold font.

2.1.2 One-Dimensional Continuous Periodic Signals

Continuous functions of a single variable that satisfy

$$f(t+T) = f(t), \quad t \in \mathbb{R} \tag{2.5}$$

are called *periodic signals*, and T defines the *period*. Such signals cannot have, in general finite $\mathcal{L}^2$ norms [Vetterli et al., 2014, p. 351]; instead, when dealing with periodic signals, it is customary to consider their energy in a single period

$$\|f\|^2 = \int_{-T/2}^{T/2} |f(t)|^2 \, dt. \tag{2.6}$$

With this caret, it is possible to identify the Hilbert space $\mathcal{L}^2([-T/2, T/2))$ of finite-energy (over one period) functions.

The circular convolution between two periodic functions $f_1(t)$ and $f_2(t)$ of period T is [Vetterli et al., 2014, p. 359]

$$(f_1 \circledast f_2)(t) = \int_{-T/2}^{T/2} f_1(\tau) f_2(t-\tau) \, d\tau = \int_{-T/2}^{T/2} f_1(t-\tau) f_2(\tau) \, d\tau. \tag{2.7}$$

The resulting function is periodic of period T, i.e. $(f_1 \circledast f_2)(t) = (f_1 \circledast f_2)(t+T)$.

2.1.3 One-Dimensional Discrete Signals

Functions of a single variable, defined on the domain of integer numbers, are also referred to as sequences. As in the continuous case, tradition motivates the association of the independent variable $l \in \mathbb{Z}$ with a time index. Signals in $\mathbb{C}^{\mathbb{Z}}$ are denoted by

$$\mathbf{f} = \begin{bmatrix} \dots & f[-2] & f[-1] & f[0] & f[1] & f[2] & \dots \end{bmatrix}^T, \tag{2.8}$$

where it is implicit that $f[l]$ corresponds to the signal sample at the lth time index and $(\cdot)^T$ denotes transposition. In many situations it is customary to consider only a finite portion of an infinite-length sequence, say of length $L \in \mathbb{N}$, as

$$\mathbf{f} = \begin{bmatrix} f[0] & f[1] & f[2] & \dots & f[L-1] \end{bmatrix}^T. \tag{2.9}$$

More formally, sequences are categorized according to the dimensionality of their support as:

- *infinite-length sequences* of the form (2.8): the domain is $\mathbb{Z}$ and they form the vector space $\mathbb{C}^{\mathbb{Z}}$ [Vetterli et al., 2014, p. 183];
- *finite-length sequences* of the form (2.9): the domain is $\{0, 1, \dots, L-1\}$ and they form the vector space $\mathbb{C}^L$ [Vetterli et al., 2014, p. 183].

Nevertheless, in this thesis finite-length sequences will not be conceived as the vector space $\mathbb{C}^L$; as a matter of fact, the circular extension of these sequences will be considered, in which the observed samples are regarded to as a single period of a periodic sequence of period L, i.e.

$$\mathbf{f} = \begin{bmatrix} \dots & f[L-1] & f[0] & f[1] & \dots & f[L-1] & f[0] & \dots \end{bmatrix}^T. \tag{2.10}$$

The standard inner product between two sequences $\mathbf{f}_1$ and $\mathbf{f}_2$ is [Vetterli et al., 2014, p. 24]

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \sum_{l=-\infty}^{\infty} f_1[l] f_2^*[l] = \mathbf{f}_2^H \mathbf{f}_1, \quad \text{if } \mathbf{f}_1, \mathbf{f}_2 \in \mathbb{C}^{\mathbb{Z}}, \tag{2.11}$$

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \sum_{l=0}^{L-1} f_1[l] f_2^*[l] = \mathbf{f}_2^H \mathbf{f}_1, \quad \text{if } \mathbf{f}_1, \mathbf{f}_2 \in \mathbb{C}^{L}, \tag{2.12}$$

where $(\cdot)^*$ and $(\cdot)^H$ denotes the operators of conjugation and of conjugate transpose, respectively. In (2.11) a slightly abused notation has been adopted, employing a vector product with infinite dimensional quantities. The inner products (2.11) and (2.12) induce the ℓ^2 (Euclidean) norms

$$\|\mathbf{f}\| = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle} = \left(\sum_{l=-\infty}^{\infty} |f[l]|^2 \right)^{1/2}, \tag{2.13}$$

$$\|\mathbf{f}\| = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle} = \left(\sum_{l=0}^{L-1} |f[l]|^2 \right)^{1/2}, \tag{2.14}$$

respectively. In the following, the attention is restricted to Hilbert spaces of *finite-energy* sequences $\ell^2(\mathbb{Z})$ and $\ell^2(\mathbb{Z}^L)$, for which the norms (2.13) and (2.14) are finite, as demonstrated in Vetterli et al. [2014, p. 185 and p. 192]. The notation $\mathbb{Z}^L$ indicates a restriction of the set $\mathbb{Z}$ to L values in the range $[0, L-1]$.

The convolution between two sequences $\mathbf{f}_1$ and $\mathbf{f}_2$ in $\mathbb{C}^{\mathbb{Z}}$ is defined as [Vetterli et al., 2014, p. 206]

$$(f_1 * f_2)[l] = \sum_{k \in \mathbb{Z}} f_1[k] f_2[l-k] = \sum_{k \in \mathbb{Z}} f_1[l-k] f_2[k]. \tag{2.15}$$

The circular convolution between two length-L sequences $\mathbf{f}_1$ and $\mathbf{f}_2$ is [Vetterli et al., 2014, p. 213]

$$(f_1 \circledast f_2)[l] = \sum_{k=0}^{L-1} f_1[k] f_2[(l-k) \bmod L] = \sum_{k=0}^{L-1} f_1[(l-k) \bmod L] f_2[k]. \tag{2.16}$$

2.1.4 Multi-Dimensional Continuous Signals

The set of complex functions of D variables, each defined on the domain of real numbers, (along with function addition and multiplication by a scalar) forms the vector space $\mathbb{C}^{\mathbb{R}^D}$. Signals in $\mathbb{C}^{\mathbb{R}^D}$ are denoted by

$$f(\mathbf{r}), \quad f : \mathbb{R}^D \to \mathbb{C}, \tag{2.17}$$

where $\mathbf{r} = [r_0, r_1, \dots r_{D-1}]^T \in \mathbb{R}^D$ denotes the index vector of the multi-dimensional signal. The standard inner product in $\mathbb{C}^{\mathbb{R}^D}$ between two multi-dimensional functions is

$$\langle f_1, f_2 \rangle = \int_{\mathbf{r} \in \mathbb{R}^D} f_1(\mathbf{r}) f_2^*(\mathbf{r}) \, d\mathbf{r}, \tag{2.18}$$

and it induces the $\mathcal{L}^2$ (Euclidean) norm

$$\|f\| = \sqrt{\langle f, f \rangle} = \left(\int_{\mathbf{r} \in \mathbb{R}^D} |f(\mathbf{r})|^2 \, d\mathbf{r} \right)^{1/2}. \tag{2.19}$$

Even with multi-dimensional signals, the energy is defined as the square of the $\mathcal{L}^2$ norm. In this thesis, the attention is restricted to the space of *finite-energy* signals $\mathcal{L}^2(\mathbb{R}^D)$, for which the $\mathcal{L}^2$ norm (2.19) is finite.

2.2 Bases Representation of Signals

The signal spaces $\mathcal{L}^2(\mathbb{R})$, $\mathcal{L}^2([-T/2, T/2])$, $\ell^2(\mathbb{Z})$ and $\ell^2(\mathbb{Z}^L)$ reviewed in Sec. 2.1 are instances of Hilbert spaces. Hilbert spaces are vectors spaces – spaces equipped with operations, such as the inner product and the norm, that induce intuitive geometric properties – characterized by the notion of *completeness*. Without entering into details, it suffices for our purposes to define the completeness as the property according to which a representative set of vectors can describe every other vector in the vector space. The interested reader can find an in-deep discussion on completeness in Vetterli et al. [2014, pp. 37-40].

The geometric intuition enabled by the Hilbert space machinery enables the introduction of the concept of signal representations, that can be regarded as coordinate systems for the Hilbert space. Representations that are not redundant are said to provide a *basis* for the Hilbert space; on the other hand, representations that are redundant are said to provide a *frame* for the Hilbert space (cfr. Sec. 2.4). In this section we review some concepts on bases, providing definitions for their specific properties.

In order to simplify the notation, here the convention introduced for the vector space $\mathbb{C}^{\mathbb{R}}$ is followed, where lowercase normal font denotes a generic element of the vector space. Despite this notational difference, the discussion holds directly also for the other vector spaces.

Given a vector space, a basis is a linearly independent set of vectors that enables a unique representation of any element of the vector space as a linear combination of basis elements. Basis are widely exploited in signal processing since they provide a mechanism to decompose a signal into a set of convenient prototype functions, from which the extraction of its salient characteristics is easier with respect to the original domain.

Given a vector space $\mathcal{V}$, the set of vectors $\Psi = \{\psi_k\}_{k\in\mathcal{K}} \subset \mathcal{V}$, $\mathcal{K}$ denoting the index set, is a basis for $\mathcal{V}$ if i) Ψ spans $\mathcal{V}$; and ii) the vectors in Ψ are linearly independent (see Strang [1998, p. 84]).

Definition 1 (Basis)**.** The set of vectors $\Psi = \{\psi_k\}_{k\in\mathcal{K}} \subset \mathcal{V}$ is a *basis* for the vector space $\mathcal{V}$ if

$$\mathcal{V} = \text{span}(\Psi) \tag{2.20}$$

and the vectors in Ψ are linearly independent.

In other words, the choice of the set Ψ has to guarantee the coverage of the vector space $\mathcal{V}$.

2.2.1 Orthonormal Bases

When the basis Ψ is formed by orthogonal and unit-norm vectors, it is called *orthonormal basis* [Strang, 1998, p. 166]; its elements satisfy

$$\langle \psi_k, \psi_i \rangle = \delta[k-i], \quad \forall k, i \in \mathcal{K}, \tag{2.21}$$

where $\delta[k]$ is the Kronecker's sequence [Vetterli et al., 2014, p. 186]

$$\delta[k] = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2.22}$$

The *analysis* of a signal into a specific orthonormal basis is accomplished by computing the *expansion coefficients* of the signal f with respect to Ψ, which are given by the projection of the element f onto each basis element ψ_k [Vetterli et al., 2014, p. 75]

$$a_k = \langle \psi_k, f \rangle. \tag{2.23}$$

Conversely, a signal can be *synthesized* from the knowledge of its expansion coefficients as the linear combination [Vetterli et al., 2014, p. 75]

$$f = \sum_{k \in \mathcal{K}} a_k \psi_k. \tag{2.24}$$

The overall representation equation can be obtained by substituting (2.23) into (2.24)

$$f = \sum_{k \in \mathcal{K}} \langle \psi_k, f \rangle \psi_k. \tag{2.25}$$

2.2.2 General Bases

When the constraint of orthogonality is relaxed (e.g. to have more freedom in choosing the basis vectors), a set of vectors can still represent any signal $f \in \mathcal{V}$. Consider two sets $\Psi = \{\psi_k\}_{k \in \mathcal{K}}$ and $\tilde{\Psi} = \{\tilde{\psi}_k\}_{k \in \mathcal{K}}$ for which the *bi-orthogonality relation* [Vetterli et al., 2014, p. 86]

$$\langle \psi_k, \tilde{\psi}_i \rangle = \delta[k - i] \tag{2.26}$$

holds. The representation equation for this pair of sets of vectors can be written as

$$f = \sum_{k \in \mathcal{K}} \langle \tilde{\psi}_k, f \rangle \psi_k = \sum_{k \in \mathcal{K}} \langle \psi_k, f \rangle \tilde{\psi}_k, \tag{2.27}$$

where it is readily apparent that the roles of ψ_k and $\tilde{\psi}_k$ can be exchanged. The two sets Ψ and $\tilde{\Psi}$ are called *bi-orthogonal bases* and they are said to form a *dual* pair. Orthonormal bases can be considered as a special case of a general basis, where the basis Ψ coincides with its dual $\tilde{\Psi}$.

We remark that the concepts of bi-orthogonal bases and dual pairs are fundamental in many signal processing applications. For instance, the almost ubiquitous *Short-time Fourier transform* (that will be introduced later in this chapter) uses a dual pair of window in its analysis and synthesis operations. Moreover, in the following of this chapter, we will generalize the notions of bi-orthogonality and dual pairs in order to introduce the properties of redundant representations of signals.

2.2.3 Analysis and Synthesis Operators

Any basis of an Hilbert space can be represented as a matrix[2] having the basis vectors as its columns [Kovacevic and Chebira, 2007, pp. 89-90]

$$\Psi = \begin{bmatrix} \psi_0 & \psi_1 & \dots \end{bmatrix}. \tag{2.28}$$

[2]Here and in the following of this chapter we adopt a slightly abused notation, according to which we use a matrix notation also for operators related to infinite-dimensional spaces: this is a widely adopted notation in the literature of signal representations.

In order for the matrix Ψ to represent a basis, it has to be non-singular, as shown in Kovacevic and Chebira [2007, p. 90]. Adopting this notation, we can define the analysis operator

$$\tilde{\Psi}^H f = a \tag{2.29}$$

and the synthesis operator

$$\Psi a = f. \tag{2.30}$$

If the basis Ψ is orthonormal, Ψ is unitary ($\Psi\Psi^H = I$, I being the identity operator) and the inverse operator is simply its adjoint, i.e. $\Psi^{-1} = \Psi^H$. In this case, the representation equation can be compactly written as [Kovacevic and Chebira, 2007, p. 90]

$$f = \sum_{k\in\mathcal{K}} \langle \psi_k, f\rangle \psi_k = \Psi\Psi^H f = \Psi^H \Psi f. \tag{2.31}$$

If the basis Ψ is not orthogonal, but it forms a bi-orthogonal pair with $\tilde{\Psi}$, we have that

$$\tilde{\Psi}^H = \Psi^{-1}; \tag{2.32}$$

the representation formula, in this case, is [Kovacevic and Chebira, 2007, p. 90]

$$f = \sum_{k\in\mathcal{K}} \langle \tilde{\psi}_k, f\rangle \psi_k = \Psi\tilde{\Psi}^H f = \tilde{\Psi}\Psi^H f = \sum_{k\in\mathcal{K}} \langle \psi_k, f\rangle \tilde{\psi}_k. \tag{2.33}$$

2.3 Fourier Transforms

In this section the Fourier transformations associated with the classes of signals presented in Section 2.1 are introduced, an interpretation of Fourier transforms in terms of expansions into specific bases is provided.

The importance of the Fourier transformations arise from the fact that complex exponentials (the bases of the Fourier transforms) are eigenfunctions of linear time-invariant systems, as described in Oppenheim et al. [1997, pp. 182-183]; i.e. applying the convolution operator to a complex exponential function gives a scalar multiple of the function itself.

2.3.1 Fourier Transform

A generic one-dimensional complex exponential function of a continuous variable $t \in \mathbb{R}$ is

$$v(t) = e^{j\omega t}, \tag{2.34}$$

where the parameter $\omega \in \mathbb{R}$ is called *angular frequency*. When the independent variable t is interpreted as time, it is customary to consider the quantity $\omega/(2\pi)$ (measured in Hz or in cycles per second).

The convolution (2.4) of $v(t)$ with an arbitrary function $h(t)$ gives [Vetterli et al., 2014, p. 360]

$$\begin{aligned}(h * v)(t) &= \int_{-\infty}^{\infty} v(t-\tau)h(\tau)\,d\tau = \int_{-\infty}^{\infty} e^{j\omega(t-\tau)}h(\tau)\,d\tau \\ &= \underbrace{e^{j\omega t}}_{v(t)} \underbrace{\int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}\,d\tau}_{H(\omega)} = v(t)H(\omega),\end{aligned} \tag{2.35}$$

where $H(\omega)$ denotes the frequency response (at angular frequency ω) of the linear time-invariant system characterized by the impulse response $h(t)$. The Fourier transform is defined as the projection of the signal $f(t)$ onto the subspaces generated by each complex exponential of the form (2.34).

Definition 2 (Fourier transform)**.** The *Fourier transform* of a signal $f \in \mathbb{C}^{\mathbb{R}}$ is

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt, \quad \omega \in \mathbb{R}. \tag{2.36}$$

The *inverse Fourier transform* of a signal $F \in \mathbb{C}^{\mathbb{R}}$ is

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t}\, d\omega, \quad t \in \mathbb{R}. \tag{2.37}$$

The notation

$$F(\omega) = \mathcal{F}_t\{f(t)\} \quad \text{and} \quad f(t) = \mathcal{F}_t^{-1}\{F(\omega)\} \tag{2.38}$$

is used to denote the direct and inverse Fourier transformations with respect to the independent variable t.

The reader interested in technical conditions for the existence of (2.36) and (2.37) is referred to specific texts on the Fourier transform, e.g. (without claims of completeness) Brémaud [2002, pp. 155-160] for a mathematical discussion and Bracewell [2000, pp. 8-9] for an engineering perspective. For our purposes, it suffices to remark the fact that the Fourier transform exists for any function $f \in \mathcal{L}^2(\mathbb{R})$ and the transformed function $F(\omega)$ is itself in $\mathcal{L}^2(\mathbb{R})$. The same fact holds also for the inverse transform.

Multi-Dimensional Signals

The Fourier transform for signals in $\mathbb{C}^{\mathbb{R}^D}$ arises as a straightforward generalization of the one-dimensional case. A complex exponential function in $\mathbb{C}^{\mathbb{R}^D}$ is

$$v(\mathbf{r}) = e^{j\langle \mathbf{k}, \mathbf{r} \rangle}, \tag{2.39}$$

where $\mathbf{k} \in \mathbb{R}^D$ plays the role of a vector of angular frequencies, each along one of the D dimensions. The multi-dimensional Fourier transform is defined as the projection of the signal $f(\mathbf{r})$ onto the subspaces generated by each complex exponential of the form (2.39).

Definition 3 (multi-dimensional Fourier transform)**.** The *Fourier transform* of a signal $f \in \mathbb{C}^{\mathbb{R}^D}$ is

$$F(\mathbf{k}) = \int_{-\infty}^{\infty} f(\mathbf{r}) e^{-j\langle \mathbf{k}, \mathbf{r} \rangle}\, d\mathbf{r}, \quad \mathbf{k} \in \mathbb{R}^D. \tag{2.40}$$

The *inverse Fourier transform* of a signal $F \in \mathbb{C}^{\mathbb{R}^D}$ is

$$f(\mathbf{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\mathbf{k}) e^{j\langle \mathbf{k}, \mathbf{r} \rangle}\, d\mathbf{k}, \quad \mathbf{r} \in \mathbb{R}^D. \tag{2.41}$$

The notation

$$F(\mathbf{k}) = \mathcal{F}_{\mathbf{r}}\{f(\mathbf{r})\} \quad \text{and} \quad f(\mathbf{r}) = \mathcal{F}_{\mathbf{r}}^{-1}\{F(\mathbf{k})\} \tag{2.42}$$

is used to denote the direct and inverse Fourier transformations with respect to $\mathbf{r}$.

2.3.2 Fourier Series

This paragraph introduces the Fourier transformation related to periodic signals in the vector space $\mathcal{L}^2([-T/2, T/2))$, including signals that arise from the circular extension of functions defined over a finite interval. The Fourier series is defined as the projection of a signal onto the subspaces generated by the eigenfunctions of the circular convolution operation. Consider a complex exponential of the form

$$v(t) = e^{j2\pi kt/T}, \tag{2.43}$$

where $k \in \mathbb{Z}$ is called *discrete frequency* and it indexes the multiples of the fundamental frequency $\omega_0 = 2\pi/T$. The circular convolution (2.7) of $v(t)$ with a function $h(t)$ gives [Vetterli et al., 2014, p. 381]

$$\begin{aligned}(h \circledast v)(t) &= \int_{-T/2}^{T/2} v(t-\tau)h(\tau)\,d\tau = \int_{-T/2}^{T/2} e^{j2\pi k(t-\tau)/T}h(\tau)\,d\tau \\ &= \underbrace{e^{j2\pi kt/T}}_{v(t)} \underbrace{\int_{-T/2}^{T/2} h(\tau)e^{-j2\pi kt/T}\,d\tau}_{H[k]},\end{aligned} \tag{2.44}$$

where $H[k]$ denotes the frequency response of the linear time-invariant system characterized by the periodic impulse response $h(t)$ at discrete frequency k. The eigenfunctions of the circular convolution operation form a countable set.

Definition 4 (Fourier series)**.** The *Fourier Series coefficient sequence* of a periodic signal $f \in \mathbb{C}^{\mathbb{R}}$ with period T is

$$F[k] = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-j(2\pi/T)kt}\,dt, \quad k \in \mathbb{Z}. \tag{2.45}$$

The *Fourier series* reconstruction for the set of coefficients $\{F[k]\}$ is

$$f(t) = \sum_{k\in\mathbb{Z}} F[k]e^{j(2\pi/T)kt}, \quad t \in [-T/2, T/2]. \tag{2.46}$$

Also in the case of the Fourier series, the conditions of existence and convergence of the integral (2.45) are not shown here. The interested reader is referred to Vetterli et al. [2014, pp. 383-385] and references therein.

Relationship between the Fourier series and the Fourier transform. The integrals (2.36) and (2.45) are equivalent if the function f takes values only in the interval $[-T/2, T/2)$ and if $\omega = 2\pi k/T$. In this case, the Fourier series coefficients are scaled samples of the Fourier transform of the same function restricted to a single period, as shown in Vetterli et al. [2014, p. 382].

Fourier Series as an Orthonormal Basis Expansion. Definition (2.45) can be interpreted as the inner product $\langle f, v_k\rangle$, where the set $\{v_k\}$ is an orthonormal basis for $\mathcal{L}^2([-T/2, T/2))$, being

$$v_k = \frac{1}{\sqrt{T}}e^{j2\pi kt/T}, \quad t \in [-T/2, T/2). \tag{2.47}$$

A proof of this statement is given in Vetterli et al. [2014, p. 384].

2.3.3 Discrete-Time Fourier Transform

The Fourier transformation related to functions of a discrete variable is introduced in this paragraph. The complex exponential sequence in $\mathbb{C}^{\mathbb{Z}}$ is

$$v[l] = e^{j\omega l}, \tag{2.48}$$

where $\omega \in \mathbb{R}$ is the angular frequency. The convolution 2.15 of $\mathbf{v} = [v[0], v[1], \ldots]$ with an arbitrary sequence $\mathbf{h}$ is [Vetterli et al., 2014, p. 217]

$$\begin{aligned}(h * v)[l] &= \sum_{k\in\mathbb{Z}} v[l-k]h[k] = \sum_{k\in\mathbb{Z}} e^{j\omega(l-k)}h[k] \\ &= \underbrace{e^{j\omega l}}_{v[l]} \underbrace{\sum_{k\in\mathbb{Z}} h[k]e^{-j\omega k}}_{H(e^{j\omega})} = v[l]H(e^{j\omega}),\end{aligned} \tag{2.49}$$

where $H(e^{j\omega})$ denotes the frequency response of a linear time-invariant system having $\mathbf{h}$ as impulse response. The discrete-time Fourier transform is defined as the projection of the signal $\mathbf{f}$ onto the subspaces generated by complex exponentials of the form (2.48).

Definition 5 (Discrete-time Fourier transform)**.** The *discrete-time Fourier transform* of a sequence $\mathbf{f}$ is

$$F(e^{j\omega}) = \sum_{l=-\infty}^{\infty} f[l]e^{-j\omega l}, \quad \omega \in \mathbb{R}. \tag{2.50}$$

The discrete-time Fourier transform exists when (2.50) converges for all $\omega \in \mathbb{R}$. $F(e^{j\omega})$ is a periodic function of period 2π. The *inverse discrete-variable Fourier transform* of a periodic function $F(e^{j\omega})$ of period 2π is

$$f[l] = \frac{1}{2\pi}\int_0^{2\pi} F(e^{j\omega})e^{j\omega l}\, d\omega, \quad l \in \mathbb{Z}. \tag{2.51}$$

The series (2.50) may fail to converge for some values of ω. However, Vetterli et al. [2014, p. 218-221] shows that (2.50) converges for all finite-energy sequences in $\ell^2(\mathbb{Z})$.

2.3.4 Discrete Fourier Transform

The instance of the Fourier transform conceived for circular extensions of finite-length sequences is introduced here. A periodic complex exponential sequence is a function of the form

$$\mathbf{v} = \begin{bmatrix} 1 & W_L^{-k} & \ldots W_L^{-(L-1)k} \end{bmatrix}^T, \tag{2.52}$$

where $W_L = e^{-j2\pi/L}$ (also referred to as *root of unity*) is an eigen-sequence for the circular convolution operation, as shown in Vetterli et al. [2014, p. 253]. Applying the circular convolution operation to the complex exponential $\mathbf{v}$ results in the same sequence scaled by the term $H[k]$, which represents the frequency response of a system having $h[n]$ as an impulse response. The discrete Fourier transform is defined as the projection of a sequence $\mathbf{f}$ onto the subspaces generated by complex exponentials of the form (2.52).

Definition 6 (Discrete Fourier transform)**.** The *discrete Fourier transform* of a length-L sequence $\mathbf{f}$ is

$$F[k] = \sum_{l=0}^{L-1} f[l]W_N^{kl}, \quad k \in \{0, \ldots, L-1\}. \tag{2.53}$$

The *inverse discrete Fourier transform* of a length-L sequence $\mathbf{F}$ is

$$f[l] = \frac{1}{L}\sum_{k=0}^{L-1} F[k]W_L^{-kl}, \quad l \in \{0, \dots, L-1\}. \tag{2.54}$$

The notation

$$F[k] = \mathcal{DFT}_L\{f[l]\} \quad \text{and} \quad f[l] = \mathcal{DFT}_L^{-1}\{F[k]\} \tag{2.55}$$

is used to denote the direct and inverse discrete Fourier transforms with respect to the independent variable l.

Upon introducing the matrix

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_L & W_L^2 & \dots & W_L^{L-1} \\ 1 & W_L^2 & W_L^4 & \dots & W_L^{2(L-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_L^{L-1} & W_L^{2(L-1)} & \dots & W_L^{(L-1)^2} \end{bmatrix}, \tag{2.56}$$

Equation (2.53) can be rewritten in matrix form as suggested by Vetterli et al. [2014, p. 254], yielding

$$\mathbf{F} = \mathbf{V}\mathbf{f}. \tag{2.57}$$

In a dual fashion, upon exploiting the orthogonality property of the roots of unity [Vetterli et al., 2014, p. 314], (2.54) can be rewritten in matrix form as [Vetterli et al., 2014, p. 254]

$$\mathbf{f} = \frac{1}{L}\mathbf{V}^H\mathbf{F}. \tag{2.58}$$

Discrete Fourier transform as an orthogonal basis expansion As stated in Vetterli et al. [2014, p. 254], Def. (2.53) can be interpreted as a set of L inner products $\langle \mathbf{f}, \mathbf{v}_k \rangle$, $k \in \{0, \dots, L-1\}$, where

$$\mathbf{v}_k = \begin{bmatrix} 1 & W_L^{-k} & \dots W_L^{-(L-1)k} \end{bmatrix}^T. \tag{2.59}$$

The basis $\{\mathbf{v}_k\}_{k=0}^{L-1}$ is orthogonal and the norm of each element is $\sqrt{L}$, as shown in Vetterli et al. [2014, p. 254]. The associated dual basis is [Vetterli et al., 2014, p. 254]

$$\tilde{\mathbf{v}}_k = \frac{1}{L}\mathbf{v}_k, \quad k \in \{0, \dots, L-1\} \tag{2.60}$$

and the relative synthesis operator yields the inverse discrete Fourier transform (2.54).

2.4 Frame Representation of Signals

When dealing with bases, the number of representative vectors of a vector space is equal to the dimension of the space itself. By adding more vectors to the representative set (thus losing the independence property), one obtains a *frame* [Vetterli et al., 2014, p. 101]. Frames enable the representation of signals in a different domain where it is easy to extract their salient characteristics, and do that in a redundant fashion. This section briefly reviews the properties of frame-based signal representations that will be used in the development of this thesis.

Definition 7 (Frame)**.** The set of vectors $\Psi = \{\psi_k\}_{k\in\mathcal{K}} \subset \mathcal{V}$ is a frame for the Hilbert space $\mathcal{V}$ if there exist two constants $0 < A < B < \infty$ such that

$$A\|f\|^2 \leq \sum_{k\in\mathcal{K}} |\langle f, \psi_k\rangle|^2 \leq B\|f\|^2, \quad \forall f \in \mathcal{V}. \tag{2.61}$$

The constants A and B are called *frame bounds*.

Given a frame $\{\psi_k\}$ of the Hilbert space $\mathcal{V}$, the synthesis operator is defined as

$$\Psi\mathbf{a} = \sum_k a_k\psi_k = f, \tag{2.62}$$

where a_k are the *expansion coefficients* of the signal f in the frame $\{\psi_k\}$ and $\mathbf{a} = \{a_k\}_{k\in\mathcal{K}}$; the adjoint of the synthesis operator is the analysis operator

$$\Psi^H f = \langle \psi_k, f\rangle. \tag{2.63}$$

The *frame operator* is defined as

$$Sf = \Psi\Psi^H f = \sum_{k\in\mathbb{Z}} \langle f, \psi_k\rangle\psi_k \tag{2.64}$$

and it is positive and invertible, as shown in Werther et al. [2005, p. 4149].

Given a frame, there exist a dual frame $\{\tilde{\psi}_k\}$ such that

$$f = \sum_{k\in\mathbb{Z}} \langle f, \tilde{\psi}_k\rangle\psi_k = \sum_{k\in\mathbb{Z}} \langle f, \psi_k\rangle\tilde{\psi}_k. \tag{2.65}$$

Given a frame $\{\psi_k\}$, the dual frame is not unique: there exist many frames that behave like duals of $\{\psi_k\}$. A possible choice for the set of vectors $\{\tilde{\psi}_k\}$ is the *canonical dual frame*

$$\{\tilde{\psi}_k\} = \{S^{-1}\psi_k\}, \tag{2.66}$$

which minimizes the norm of the expansion coefficients. Later in this thesis, more precisely in Sec. 7.5, we review an effective solution to compute the dual frame in a finite-dimensional setting.

2.4.1 Gabor Frames

Gabor frames are one class of frames of widespread use in the signal processing community. The constituting elements of a Gabor frame are

- a prototype function $\psi \in \mathcal{L}^2$;
- a time parameter α; and,
- a frequency parameter β.

A Gabor frame is defined in the literature, e.g. in Werther et al. [2005], as a sequence of translations in time and modulations in frequency applied to the prototype function ψ. Introducing the translation and modulation operators

$$T_{n\alpha}\psi = \psi_{t-n\alpha}, \quad n \in \mathbb{Z} \tag{2.67}$$

$$M_{m\beta}\psi = \psi e^{jm\beta t}, \quad m = \mathbb{Z}, \tag{2.68}$$

we can compactly denote a Gabor frame with elements (ψ, α, β) as the set of functions

$$\{T_{n\alpha} M_{m\beta} \psi\}_{m,n\in\mathbb{Z}}. \tag{2.69}$$

A widely adopted choice for the prototype function ψ is the dilated Gaussian function, which in $\mathbb{C}^{\mathbb{R}}$ is defined as

$$\psi(t) = e^{-\pi t^2/\sigma^2}, \quad \sigma \in \mathbb{R}. \tag{2.70}$$

Its widespread use is motivated by the fact that the Gaussian function exhibit the best trade-off between time-frequency localization. In Chap. 7 we will discuss with more details the choice of the Gaussian function.

Gabor frames are of fundamental importance in this thesis. As a matter of fact, in Sec. 4.4 we show that Gabor frames can be used to build a physically motivated and rigorous representation for acoustic fields. Moreover, in Chap. 7 we exploit the properties of Gabor frames in order to define a new transformation for acoustic fields that allows to represent their information in a specific domain, where their salient characteristics are readily evident.

2.5 Local Fourier Transform

The Fourier transformations reviewed in Sec. 2.3 are fundamental tools for the global analysis of a signals, in the sense that all the time-distributed information encoded by the signal are integrated to provide a single information. However, in many situations occurring in practical scenarios, we are interested in local properties of a signal, i.e. in following the evolution of its frequency content along with time. For the sake of an example, consider a time-domain signal $f(t)$ encoding a musical melody. A melody can be conceived, in its simplest form, as a succession of notes, i.e. events characterized by *pitch* and *duration*; thus melody is a simple example of a time-domain signal whose frequency content varies with time. The Fourier transform $F(\omega)$ is not suitable to capture this temporal evolution, as it considers the whole duration of the signal $f(t)$ itself.

For the purpose of processing this class of signals, several transforms have been introduced in the literature, originating from the work in Gabor [1946] on time-frequency localization. These transformations are known under many names, i.e. *windowed Fourier transform*, as in Brémaud [2002, Chap. D1]; *short-term Fourier transform* [Allen, 1977]; *short-time Fourier transform* [Vetterli and Kovacevic, 1995, Sec. 2.6.3], according to the localizing function that is adopted; in this thesis we adopt the nomenclature *local Fourier transform*, as in Vetterli et al. [2014, Chap. 7], in order to highlight its local nature, in contrast to the global nature of the Fourier transform.

2.5.1 Local Continuous Fourier Transform

We define a real function $\psi(t) \in \mathbb{R}^{\mathbb{R}}$, whose support is confined to a small interval about zero; moreover, $\psi(t)$ is assumed to exhibit even symmetry (i.e. $\psi(t) = \psi(-t)$). Examples of possible function $\psi(t)$ are the rectangular window of width σ

$$\psi(t) = \begin{cases} 1, & \text{if } -\sigma/2 \le t \le \sigma/2 \\ 0, & \text{otherwise,} \end{cases} \tag{2.71}$$

and the dilated Gaussian window

$$\psi(t) = e^{-t^2/\sigma^2}, \quad \sigma \in \mathbb{R}^+. \tag{2.72}$$

The local information of a signal $f(t) \in \mathbb{C}^{\mathbb{R}}$ can be retrieved by projecting it onto the space of time-localized complex exponential functions

$$v_\tau(t) = \psi(t-\tau)e^{j\omega t}, \quad \tau \in \mathbb{R}, \quad \omega \in \mathbb{R}, \tag{2.73}$$

where τ denotes the temporal centroid of the window. The Local Fourier transform is then introduced starting from the windowed complex exponential function, and it is defined as the inner product between f and $v_\tau(t)$.

Definition 8 (Local Fourier transform)**.** The *local Fourier transform* of a signal $f(t) \in \mathbb{C}^{\mathbb{R}}$ is

$$F(\omega, \tau) = \langle f, v_\tau \rangle = \int_{t\in\mathbb{R}} f(t) v_\tau^H(t)\, dt, \quad \omega, \tau \in \mathbb{R}. \tag{2.74}$$

The *inverse local Fourier transform* of $F(\omega, \tau)$ is

$$f(t) = \frac{1}{2\pi} \int_{\omega\in\mathbb{R}} \int_{\tau\in\mathbb{R}} F(\omega, \tau) v_\tau(t)\, d\tau\ d\omega. \tag{2.75}$$

It can be proved (the interested reader is referred to Brémaud [2002, pp. 178-182] for a proof) that the inverse local Fourier transform holds under the following assumptions

1. $\int_{\mathbb{R}} |\psi(t)|\, dt < \infty$, i.e. $\psi(t) \in \mathcal{L}^1(\mathbb{R})$)
2. $\int_{\mathbb{R}} |\psi(t)|^2\, dt = 1$.

Assumption 2 is just a convention in order for the window function to have unitary energy: more generally, it is sufficient to assume $\psi(t) \in \mathcal{L}^2(\mathbb{R})$.

The local Fourier transform provides a highly redundant representation of the signal $f(t)$. Indeed, it maps the one-dimensional signal $f(t)$ to a two-dimensional signal $F(\omega, \tau)$. This mapping is based on time-frequency functions $v_\tau(t)$ that exhibit the most of their energy in the vicinity of the point (τ, ω) in the time-frequency plane.

2.5.2 Local Discrete Fourier Transform

The one-dimensional local Discrete Fourier transform for finite-length discrete signals is introduced as the discrete counterpart of the local Fourier transform introduced in the preceding paragraph. A discrete real sequence $\psi[l] \in \mathbb{R}^L$ is defined, whose support is included in a small interval about $l = \tau$, $\tau \in \mathbb{Z}$. The local information of a signal $f[l] \in \mathbb{C}^L$ around time index $n = \tau$ is thus contained in the discrete signal

$$f[l]\psi[l-\tau], \quad l = 0, \ldots, L-1. \tag{2.76}$$

The Local Discrete Fourier transform is introduced starting from the windowed discrete complex exponential function

$$v_\tau[l] = \psi[l-\tau] W_L^{-kl}, \quad \tau \in \{0, \ldots, L-1\}, \quad k = \{0, \ldots, L-1\}, \tag{2.77}$$

where W_L is the root of unity introduced in (2.52).

Definition 9 (Local discrete Fourier transform)**.** The *local discrete Fourier transform* of a signal $\mathbf{f} \in \mathbb{C}^L$ is

$$F[\tau, k] = \sum_{l=0}^{L-1} f[l]\psi[l-\tau] W_L^{-kl}, \quad \tau \in \{0, \ldots, L-1\}, \quad k = \{0, \ldots, L-1\}. \tag{2.78}$$

The *inverse local discrete Fourier transform* of $F[\tau, k]$ is

$$f[l] = \frac{1}{L} \sum_{\tau=0}^{L-1} \sum_{k=0}^{L-1} F[\tau, k] W_L^{kl}. \tag{2.79}$$

Table 2.1: Summary of various Fourier transform instances. Adapted from [Vetterli et al., 2014, p. 401, Tab. 4.4].

Transform	Analysis/Synthesis	Characteristics
Fourier transform	$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt$	
(Def. 2)	$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega$	
Fourier series	$F[k] = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-j(2\pi/T)kt}\,dt$	Dual with DTFT
(Def. 4)	$f(t) = \sum_{k\in\mathbb{Z}} F[k]e^{j(2\pi/T)kt}$	$f(t+T) = f(t)$
Discrete-time	$F(e^{j\omega}) = \sum_{l=-\infty}^{\infty} f[l]e^{-j\omega l}$	Dual with Fourier Series
Fourier transform (Def. 5)	$f[l] = \frac{1}{2\pi}\int_{0}^{2\pi} F(e^{j\omega})e^{j\omega l}\,d\omega$	$F(e^{j(\omega+2\pi)}) = F(e^{j\omega})$
Discrete Fourier transform	$F[k] = \sum_{l=0}^{L-1} f[l]W_L^{kl}$	
(Def. 6)	$f[l] = \frac{1}{L}\sum_{k=0}^{L-1} F[k]W_L^{-kl}$	
Local Fourier transform	$F(\omega,\tau) = \int_{-\infty}^{\infty} f(t)v_\tau^*(t)\,dt$	
(Def. 8)	$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega,\tau)v_\tau(t)\,d\tau\,d\omega$	
Local Discrete Fourier	$F[\tau,k] = \sum_{l=0}^{L-1} f[l]v_\tau^*[l]$	
transform (Def. 9)	$f[l] = \frac{1}{L}\sum_{\tau=0}^{L-1}\sum_{k=0}^{L-1} F[\tau,k]v_\tau[l]$	

2.6 Summary

- A unified formalism based on the notion of Hilbert spaces is introduced to deal with different classes of signals
- Bases and frames are presented as tools that enable the representation of signals in a domain where signal characteristics are more evident.
- Fourier bases (complex exponentials) emerge as one of the most useful signal representations; Fourier transformations are defined as the projection of a signal f onto the subspaces generated by complex exponential signals.
- Different instances of the Fourier transformations arise when considering complex exponentials signals in different vector spaces: a summary is provided in Tab. 2.1.
- Multi-dimensional instances of the Fourier transformations arise as extensions of the one-dimensional cases.
- Local instances of the Fourier transforms are introduced by windowing the complex exponential signals.
- Gabor frames are introduced as an useful instance of general frames.

This page intentionally left blank.

Chapter 3

Fundamentals of Acoustic Theory

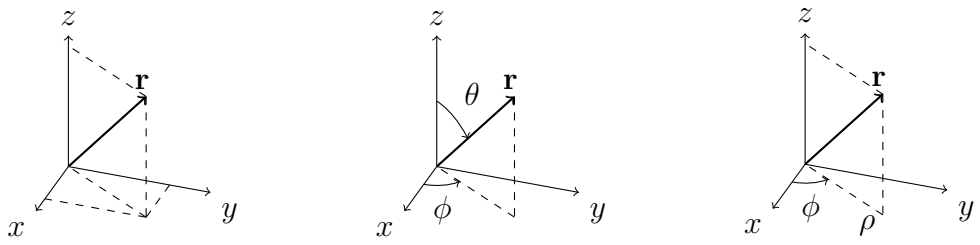
This chapter reviews the physical laws governing sound generation and propagation. In this chapter sound is described by a scalar function, the *acoustic field*. An acoustic field is a real-valued scalar function (or distribution) whose domain is the union of time and space regions; in mathematical terms, the function

$$p(\mathbf{r}, t), \quad \mathbf{r} \in \mathbb{R}^3, \quad t \in \mathbb{R} \tag{3.1}$$

is the general form of an acoustic field, with $\mathbf{r}$ and t denoting the spatial coordinates and time, respectively. Acoustic fields (unlike, e.g., electromagnetic fields and all vector fields) satisfy the property of *invariance* under a transformation of space coordinates, meaning that the value of the field at a point in space is independent on the coordinate system adopted to represent that point, as discussed in Morse and Feshbach [1953, p. 4]. What varies is the functional form of the mathematical expression for the field. In other words, $p(\mathbf{r}, t)$ encodes the same numerical value regardless of the specific coordinate system adopted to represent the spatial variable $\mathbf{r}$ (e.g. Cartesian coordinates, spherical coordinates, etc.). This consideration motivates the use of different *representations* for the acoustic field in Chap. 4, each linked to a specific choice of the reference frame.

Sections 3.1-3.6 review the fundamentals of *wave acoustics*, i.e. the study of sound propagation in the form of waves. This paradigm, on which an excellent review can be found in Williams [1999], is the most widely adopted in the acoustics signal processing literature and serves as a foundation for many applications. On the other hand, Sec. 3.7-3.8 review two fundamentals approaches to geometrical acoustics, the first based on *acoustic rays*, while the latter based on *acoustic beams*. We remark that, at our knowledge, the physical model of acoustic beams has not been considered so far in the literature on acoustic signal processing, hence we consider it as a novel contribution provided by this thesis.

Before proceeding with the discussion, we introduce the notation used in the rest of this thesis to denote spatial coordinate systems. We will consider 3D spatial coordinates, with time being the fourth dimension. Consider a position vector $\mathbf{r}$ in a three-dimensional space. In Cartesian coordinates the position vector $\mathbf{r} = [x, y, z]^T$ denotes the three spatial coordinates in a right-handed orthogonal coordinate system, as shown in Fig. 3.1a. Unit vectors in the three spatial directions are denoted by $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$. On the other hand, a point in a spherical reference frame is represented by its distance r from the origin, its azimuth ϕ and its co-elevation θ, as shown in Fig. 3.1b. Spherical coordinates are related



(a) Cartesian. **(b)** Spherical. **(c)** Cylindrical.

Figure 3.1: Three-Dimensional coordinate systems.

to Cartesian coordinates via the relationships

$$\begin{aligned} x &= r\sin(\theta)\cos(\phi) \\ y &= r\sin(\theta)\sin(\phi) \\ z &= r\cos(\theta) \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} r &= \sqrt{x^2+y^2+z^2} \\ \phi &= \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right) = \arcsin\left(\frac{y}{\sqrt{x^2+y^2}}\right) \\ \theta &= \arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right). \end{aligned} \tag{3.3}$$

Finally, a point in a cylindrical reference frame is represented by its radial distance ρ, its azimuth ϕ and its height z, as shown in Fig. 3.1c. Cylindrical coordinates are related to Cartesian coordinates via the relationships

$$\begin{aligned} x &= \rho\cos(\phi) \\ y &= \rho\sin(\phi) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \rho &= \sqrt{x^2+y^2} \\ \phi &= \arctan\left(\frac{y}{x}\right). \end{aligned} \tag{3.5}$$

3.1 The Laplace Operator

A key property of an acoustic field involves its second order derivatives with respect to the spatial coordinates. This property measures the curvature of the function $p(\mathbf{r}, t)$ and its mathematical description is provided by the *Laplace operator* [Morse and Feshbach, 1953, p. 7], denoted by the symbol ∇^2 and defined as the divergence of the gradient operator

$$\nabla^2 = \nabla \cdot (\nabla). \tag{3.6}$$

The Laplace operator takes different forms in different coordinate systems. Equations (3.7)-(3.9) are the functional forms of the Laplace operator expressed in the coordinate systems considered in this thesis. These expressions can be obtained starting from the definition of

the gradient and divergence operators in different coordinate systems, and then applying the coordinate transformations listed in the previous section.

$$\text{Cartesian} \quad \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{3.7}$$

$$\text{Spherical} \quad \nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial \theta}\left(\sin(\theta)\frac{\partial}{\partial \theta}\right) + \frac{1}{r^2\sin^2(\theta)}\frac{\partial}{\partial \phi^2} \tag{3.8}$$

$$\text{Cylindrical} \quad \nabla^2 = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}. \tag{3.9}$$

3.2 The Wave Equations

3.2.1 The Homogeneous Wave Equation

In the linear regime, the acoustic field can be described by small-amplitude variations of the pressure $p(\mathbf{r}, t)$ as a function of the space variable $\mathbf{r}$ and time variable t. Considering a volume free of sources, the function $p(\mathbf{r}, t)$, in order to be a valid acoustic field, must satisfy the *homogeneous wave equation*

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0, \tag{3.10}$$

where c denotes the speed of sound in air (expressed in $\mathrm{m\,s^{-1}}$). We refer the interested reader to Morse and Ingard [1986, Chap. 6] or to Kinsler et al. [2000, Chap. 5] for a derivation of (3.10) starting from basic physical principles. Equation (3.10) governs the propagation of acoustic waves in both space and time domains.

In many contexts it is useful to consider acoustic propagation as a function of space and temporal frequency. This is the case, for instance, of an acoustic field generated by a source in harmonic motion. In this case, the dependence on time can be expressed as $p(\mathbf{r}, t) = P(\mathbf{r}, \omega)e^{j\omega t}$, being $P(\mathbf{r}, \omega) = \mathcal{F}_t\{p(\mathbf{r}, t)\}$. The equation equivalent to (3.10) in the temporal-frequency domain can be obtained starting from the Fourier transform (Def. 2) of (3.10) with respect to time

$$\mathcal{F}_t\left\{\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2}\right\} = 0, \tag{3.11}$$

which can be simplified by exploiting the linearity property of the Fourier transform [Vetterli et al., 2014, p. 365] to yield

$$\nabla^2 \mathcal{F}_t\{p(\mathbf{r}, t)\} - \frac{1}{c^2}\mathcal{F}_t\left\{\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2}\right\} = 0. \tag{3.12}$$

Upon substituting the definition of $P(\mathbf{r}, \omega)$ and applying the differentiation property of the Fourier transform [Vetterli et al., 2014, p. 367], it results

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = 0. \tag{3.13}$$

Equation (3.13) is known as *Helmholtz equation*, as described in Williams [1999, p. 18].

3.2.2 The Inhomogeneous Wave Equation

If the volume under consideration is not free of sources, the wave equation (3.10) can not be used to describe the acoustic field in the volume. As a matter of fact, the wave equation must include a source excitation term, describing the energy emitted by the sound source. This is accomplished by replacing the right-hand side of (3.10) with a suitable source term depending on the volume flow rate per unit volume $q(\mathbf{r}, t)$, yielding the *inhomogeneous wave equation*

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -\frac{\partial q(\mathbf{r}, t)}{\partial t}. \tag{3.14}$$

We refer the interested reader to Kinsler et al. [2000, Sec. 5.15] for a derivation of (3.14).

Similarly with what has been done for the homogeneous wave equation, we consider also the formulation of the inhomogeneous wave equation in the temporal-frequency domain. As in the homogeneous case, the *inhomogeneous Helmholtz equation* can be derived by taking the Fourier transform of (3.14) with respect to time, yielding

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = -j\omega Q(\mathbf{r}, \omega). \tag{3.15}$$

3.3 Solutions to the Homogeneous Wave Equations

In this section we review acoustic fields that emerge as solutions to the homogeneous wave equation presented in the previous section. Each solution emerges from the choice of the actual reference system in use (Cartesian, spherical, cylindrical). Time-harmonic dependence for the acoustic fields is assumed; this assumption allows us to consider the equivalent (but easier) problem of seeking solutions to the homogeneous Helmholtz equation (3.13).

3.3.1 Plane Waves

We start by reviewing solutions to the homogeneous Helmholtz equation (3.13) in a Cartesian reference frame. Candidate solutions are in the form of complex exponential functions, as assumed by Williams [1999, pp. 21-23] and Kinsler et al. [2000, pp. 121-124]. Therefore we can write

$$P(\mathbf{r}, \omega) = e^{j\langle \mathbf{k}, \mathbf{r} \rangle}, \tag{3.16}$$

where $\mathbf{k} = [k_x, k_y, k_z]$ is called the *wavenumber vector*. The function (3.16) is a solution to (3.13) as long as $\mathbf{k}$ satisfies the *dispersion relation*

$$\|\mathbf{k}\|^2 = \left(\frac{\omega}{c}\right)^2. \tag{3.17}$$

The dispersion relation establishes the dependency of the wavenumber vector on frequency. In the following we consider two cases:

- the case of propagating plane waves, where $\mathbf{k} \in \mathbb{R}^3$; and,
- the case of evanescent plane waves, where one component of $\mathbf{k}$ assumes imaginary values, while the other two components are real.

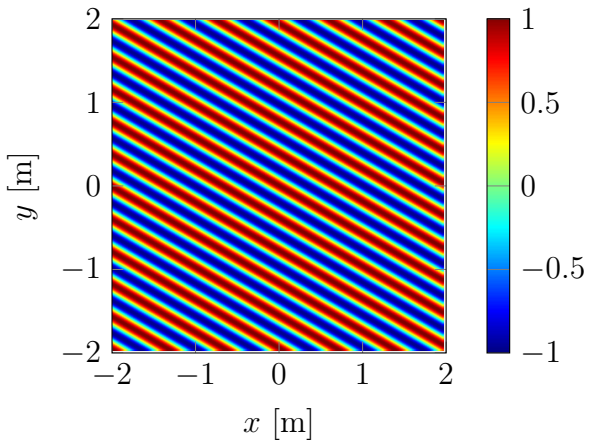


Figure 3.2: Acoustic field of a propagating plane wave ($\Re\{P(\mathbf{r},\omega)\}$). The wavenumber vector is $\mathbf{k} = (\omega/c)[1/2, \sqrt{3}/2, 0]^T$, $\omega = 2\pi f$, $f = 1\,\text{kHz}$, $c = 340\,\text{m}\,\text{s}^{-1}$. The acoustic field is shown on the plane $z = 0\,\text{m}$.

Propagating Plane Waves. Consider the function (3.16) and restrict $\mathbf{k}$ to be in $\mathbb{R}^3$. The *locus* of points where the phase of the function is constant, usually referred to as *wavefront*, arises as the solution to

$$\angle P(\mathbf{r},\omega) = \langle \mathbf{k}, \mathbf{r} \rangle = C, \tag{3.18}$$

being $C \in \mathbb{R}$ a constant. It is evident from (3.18) that the *locus* of points at constant phase is a plane in $\mathbb{R}^3$, orthogonal to the vector $\mathbf{k}$. This basic reasoning motivates the name *plane wave* assigned to the function (3.16). The unit vector in the direction of $\mathbf{k}$, $\hat{\mathbf{k}} = \mathbf{k}/\|\mathbf{k}\|$, is identified as the Direction of Arrival of the plane wave [1]. Figure 3.2 shows the real part of the acoustic field $P(\mathbf{r},\omega)$ of a propagating plane wave with $\mathbf{k} = \omega/c[1/2, \sqrt{3}/2, 0]^T$, $\omega = 2\pi f$, $f = 1\,\text{kHz}$, $c = 340\,\text{m}\,\text{s}^{-1}$. The acoustic field is shown on the plane $z = 0\,\text{m}$.

The corresponding solution to the homogeneous wave equation (3.10) is found by adding the time-harmonic dependence to (3.16)

$$p(\mathbf{r},t) = P(\mathbf{r},\omega)e^{j\omega t} = e^{j(\langle \mathbf{k},\mathbf{r}\rangle + \omega t)}. \tag{3.19}$$

Wavelength. Along with the direction of arrival, a plane wave is characterized by its *wavelength* λ, defined as the spatial distance traveled by a plane of constant phase in a temporal period $T = 2\pi/\omega$. Consider a plane of constant phase displaced by $\delta\mathbf{r}$ in a time interval $\delta t = 2\pi/\omega$

$$\angle p(\mathbf{r},t) = \angle p(\mathbf{r} + \delta\mathbf{r}, t + \delta t), \tag{3.20}$$

which can be rewritten as

$$\langle \mathbf{r}, \mathbf{k} \rangle + \omega t = \langle \mathbf{r} + \delta\mathbf{r}, \mathbf{k} \rangle + \omega(t + \delta t). \tag{3.21}$$

Exploiting the distributivity property of the inner product one can write

$$\langle \mathbf{r}, \mathbf{k} \rangle + \omega t = \langle \mathbf{r}, \mathbf{k} \rangle + \langle \delta\mathbf{r}, \mathbf{k} \rangle + \omega t + \omega\delta t, \tag{3.22}$$

[1] We remark that this is an arbitrary convention. Some authors prefer to identify the wavenumber vector with the direction of propagation of the plane wave. In this thesis we adhere to the convention of $\hat{\mathbf{k}}$ identifying the DoA. Following this convention, we identify the direction of propagation as $-\hat{\mathbf{k}}$.

which can be solved for $\delta\mathbf{r}$ after the elimination of common terms

$$\langle \delta\mathbf{r}, \mathbf{k} \rangle = -\omega\delta t. \tag{3.23}$$

The wavelength λ can be defined as the norm of the displacement vector $\delta\mathbf{r}$ in (3.23), i.e. $\lambda = \|\delta\mathbf{r}\|$. We define the angle α between the vectors $\delta\mathbf{r}$ and $\mathbf{k}$; the considered setting dictates $\alpha = \pi$ (we recall that $\hat{\mathbf{k}}$ denotes the DoA of the plane wave, which is opposite to the direction of propagation). The inner product in (3.23) can be expanded in terms of the norms of the two vectors, and then, by solving for λ, it results

$$\lambda = \frac{2\pi}{\|\mathbf{k}\|}. \tag{3.24}$$

Spatial Frequency. The magnitude of the wavenumber vector $\mathbf{k}$ expresses the number of cycles per unit length that the plane wave exhibits along the DoA. Thus, it can be immediately seen that $\mathbf{k}$ plays the role of a spatial frequency variable, in analogy with the role of ω as a temporal frequency (which expresses the number of cycles per time unit).

Evanescent Plane Waves. The wavenumber is not constrained to be real in order to satisfy the dispersion relation (3.17). Indeed, one can relax the constraint $\mathbf{k} \in \mathbb{R}^3$ by letting two components of $\mathbf{k}$ being real, i.e. $k_x, k_y \in \mathbb{R}$, while the third component being imaginary, i.e. $k_z = j\zeta$, $\zeta \in \mathbb{R}$. Thus, the plane-wave function (3.16) can be written as

$$p(\mathbf{r}, t) = e^{j\omega t} e^{j(k_x x + k_y y)} e^{-\zeta z}. \tag{3.25}$$

With reference to (3.25), it is customary to ensure that the complex exponential function $e^{j\langle \mathbf{k}, \mathbf{r} \rangle}$ makes physical sense: i.e. we discard functions that gain energy as they propagate, as these functions have no physical sense. This is equivalent to constraining ζ to be positive, i.e. $\zeta \in \mathbb{R}^+$. The class of functions satisfying the above described requirements is called *evanescent waves* and consists of plane wave functions whose amplitude decays exponentially with space. These functions are still solution to (3.2) but they do not propagate.

In order to derive a condition on the components of the wavenumber vector that enables the generation of evanescent waves, the dispersion relation (3.17) is rewritten as

$$k_z^2 = \left(\frac{\omega}{c}\right)^2 - k_x^2 - k_y^2 \quad \text{and} \quad k_z = \sqrt{\left(\frac{\omega}{c}\right)^2 - k_x^2 - k_y^2}, \tag{3.26}$$

from which it is evident that k_z is imaginary when $k_x^2 + k_y^2 > (\omega/c)^2$. Figure 3.3 shows the real part of the acoustic field of an evanescent plane wave with $\mathbf{k} = \omega/c[\sqrt{1.01}, 0, j0.1]^T$, $\omega = 2\pi f$, $f = 1\,\text{kHz}$, $c = 340\,\text{m}\,\text{s}^{-1}$. The acoustic field is shown on the plane $y = 0\,\text{m}$.

3.3.2 Spherical Waves

Another elementary solution to the homogeneous Helmholtz equation is the *spherical wave*, originated by a sound source of infinitesimal dimension, usually referred to as *point source* or *acoustic monopole*, placed outside the spatial region of interest. The acoustic field of a point source is an *outward* spherical wave of the form presented in Williams [1999, p. 198], i.e.

$$P(\mathbf{r}, \omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|}, \tag{3.27}$$

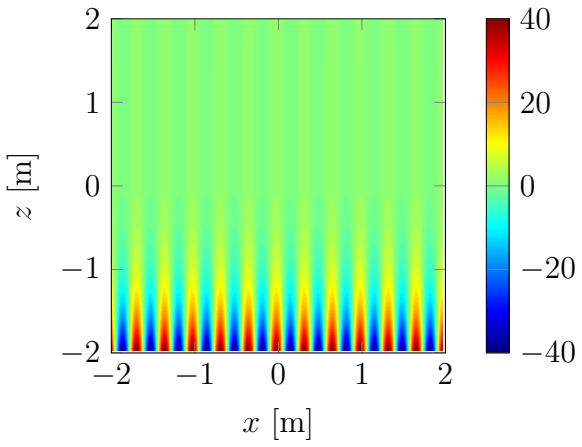


Figure 3.3: Acoustic field of an evanescent plane wave (real part). The wavenumber vector is $\mathbf{k} = (\omega/c)[\sqrt{1.01}, 0, j0.1]^T$, $\omega = 2\pi f$, $f = 1\,\text{kHz}$, $c = 340\,\text{m}\,\text{s}^{-1}$. The acoustic field is shown on the plane $y = 0\,\text{m}$.

where $\mathbf{r}'$ denotes the spatial position of the point source. The *loci* of constant phase of (3.27) are the surfaces $(\omega/c)\|\mathbf{r} - \mathbf{r}'\| = 2\pi C$, $C \in \mathbb{Z}$, which are concentric spheres centered at $\mathbf{r}'$ and separated by the wavelength $\lambda = 2\pi/(\omega/c)$. Spherical waves of the form (3.27) are outward spherical waves, in the sense that the wavefronts propagate from $\mathbf{r}'$ to ∞. On the other hand, *inward* spherical waves generate acoustic fields of the form

$$P(\mathbf{r}, \omega) = \frac{e^{j\frac{\omega}{c}\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r} - \mathbf{r}'\|}, \tag{3.28}$$

which exhibit spherical wavefronts propagating from infinity to an infinitesimal sink of sound at $\mathbf{r}'$.

In many situations, it is convenient to deal with acoustic fields generated by spherical waves in a coordinate system different from the Cartesian one. The natural choice is the spherical reference frame. This paragraph reviews solutions to the homogeneous Helmholtz equation in a spherical reference frame. In this setting, the Laplace operator is (3.8) and the homogeneous Helmholtz equation (3.13) reads

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial P(\mathbf{r},\omega)}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial P(\mathbf{r},\omega)}{\partial\theta}\right) + \\ + \frac{1}{r^2\sin^2(\theta)}\frac{\partial^2 P(\mathbf{r},\omega)}{\partial\phi^2} + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r},\omega) = 0. \tag{3.29}$$

Solutions to (3.29) are obtained by separation of variables. The acoustic field $P(\mathbf{r}, \omega)$ is assumed in the factored form [Morse and Ingard, 1986, p. 333]

$$P(\mathbf{r}, \omega) = R(r)\Theta(\theta)\Phi(\phi), \tag{3.30}$$

hence (3.29) can be rewritten as

$$\frac{\Theta(\theta)\Phi(\phi)}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right) + \frac{R(r)\Phi(\phi)}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta(\theta)}{\partial\theta}\right) + \\ + \frac{R(r)\Theta(\theta)}{r^2\sin^2(\theta)}\frac{\partial^2\Phi(\phi)}{\partial\phi^2} + \left(\frac{\omega}{c}\right)^2 R(r)\Theta(\theta)\Phi(\phi) = 0. \tag{3.31}$$

After multiplication by $r^2/(R(r)\Theta(\theta)\Phi(\phi))$, (3.31) reads

$$\frac{1}{R(r)}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right)+\frac{1}{\Theta(\theta)\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta(\theta)}{\partial\theta}\right)+ \\ +\frac{1}{\Phi(\phi)\sin^2(\theta)}\frac{\partial^2\Phi(\phi)}{\partial\phi^2}+\left(\frac{\omega}{c}\right)^2 r^2=0. \quad (3.32)$$

Dependence on Azimuth. Consider first the dependence of (3.32) on the azimuth angle ϕ. The only term in (3.32) depending on ϕ is the third term, while all the others can be regarded to as constants. After a multiplication by $\sin^2(\theta)$, it results that the function $\Phi(\phi)$ must be a solution to

$$\frac{1}{\Phi(\phi)}\frac{\partial^2\Phi(\phi)}{\partial\phi^2}=-m^2, \quad (3.33)$$

where m is a constant. According to Morse and Ingard [1986, p. 333] and Williams [1999, p. 185], the solution to (3.33) is

$$\Phi(\phi)=\Phi_1 e^{jm\phi} \quad (3.34)$$

with the constraint $m\in\mathbb{Z}$ to avoid discontinuities at ϕ equal to integer multiples of 2π.

Dependence on Co-Elevation. Upon substituting (3.34) into (3.32), the latter can be rewritten as

$$\frac{1}{R(r)}\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right)+\frac{1}{\Theta(\theta)\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta(\theta)}{\partial\theta}\right)-\frac{m^2}{\sin^2(\theta)}+\left(\frac{\omega}{c}\right)^2 r^2=0. \quad (3.35)$$

The first and the last terms in (3.35) do not depend on θ, hence they can be considered constant. Upon defining this constant as $l(l+1)$[2], one gets

$$\frac{1}{\Theta(\theta)\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta(\theta)}{\partial\theta}\right)-\frac{m^2}{\sin^2(\theta)}=-l(l+1). \quad (3.36)$$

Upon multiplying by $\Theta(\theta)$ and rearranging terms, (3.36) can be rewritten as

$$\frac{1}{\Theta(\theta)\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial\Theta(\theta)}{\partial\theta}\right)+\left(l(l+1)-\frac{m^2}{\sin^2(\theta)}\right)=0. \quad (3.37)$$

With the substitution $a=\cos(\theta)$, (3.37) can be rewritten in the form of the *Associated Legendre Equation* [Olver, 2010, p. 352]

$$\frac{\partial}{\partial a}\left((1-a^2)\frac{\partial\Theta(a)}{\partial a}\right)+\left(l(l+1)-\frac{m^2}{1-a^2}\right)\Theta(a)=0, \quad (3.38)$$

whose solution is [Williams, 1999, p. 185]

$$\Theta(a)=\mathsf{P}_l^m(a), \quad a=\cos(\theta), \quad (3.39)$$

$\mathsf{P}_l^m(a)$ being the *Associated Legendre Polynomial* [Olver, 2010, p. 360] of integer degree l and order m ($l\in\mathbb{N}$ and $m=-l,\ldots,l$).

[2]The choice of the constant as $l(l+1)$ is dictated by the availability of a closed-form solution. For other choices no solution does exist [Arfken and Weber, 2005, Chap. 12].

Spherical Harmonics. The functions (3.33) and (3.39) are usually combined, as in Morse and Ingard [1986, p. 335], into the *spherical harmonic* function [Ahrens, 2012, p. 27]

$$Y_l^m(\theta,\phi) = (-1)^m \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(n+|m|)!}} e^{jm\phi} \mathsf{P}_l^{|m|}(\cos(\theta)). \tag{3.40}$$

We remark that the angular dependence of a generic solution to the Helmholtz equation in spherical coordinates, as encoded by the spherical harmonic functions, is independent on temporal frequency.

The advantage of (3.40) with respect to other definitions for the spherical harmonics (e.g. the ones in Olver [2010, p. 378, Eq. 14-30.1] and in Williams [1999, p. 186, Eq. 6.20]) is that it enables to directly handle the case $m < 0$ and, moreover, the conjugation takes the convenient form [Ahrens, 2012, p. 27]

$$(Y_l^m(\theta,\phi))^* = Y_l^{-m}(\theta,\phi). \tag{3.41}$$

Spherical harmonics of the form (3.40) enjoy the properties of orthogonality and normalization [Olver, 2010, p. 336]

$$\int_0^{2\pi}\int_0^{\pi} Y_l^{-m}(\theta,\phi) Y_{l'}^{m'}(\theta,\phi)\sin(\theta)\,d\theta\,d\phi = \delta[l-l']\delta[m-m'], \tag{3.42}$$

where $\delta[\cdot]$ denotes the Kronecker delta sequence (2.22).

Dependence on Radius. The radial dependence is obtained by substituting (3.34) and (3.39) into (3.32), yielding [Morse and Ingard, 1986, p. 335]

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial R(r)}{\partial r}\right) + \left(\left(\frac{\omega}{c}\right)^2 r^2 - l(l+1)\right) R(r) = 0. \tag{3.43}$$

As illustrated in Williams [1999, pp. 185-186], the solution to (3.43) can be expressed in terms of *spherical Bessel* functions

$$R(r) = R_1 j_l((\omega/c)r) + R_2 y_l((\omega/c)r), \tag{3.44}$$

$j_l(\cdot)$ being the spherical Bessel function of first kind [Olver, 2010, p. 262, Eq. 10.47.3], $y_l(\cdot)$ denotes the spherical Bessel function of second kind [Olver, 2010, p. 262, Eq. 10.47.4] and R_1, R_2 are constants. Alternatively, the solution to (3.43) can be written in terms of *spherical Hankel* functions

$$R(r) = R_3 h_l^{(1)}((\omega/c)r) + R_4 h_l^{(2)}((\omega/c)r), \tag{3.45}$$

where $h_l^{(1)}(\cdot)$ and $h_l^{(2)}(\cdot)$ are spherical Hankel function of first and second kind, respectively [Olver, 2010, p. 262, Eqs. 10.47.5-6]. The coefficients R_1, R_2, R_3, R_4 are usually chosen in order to keep or discard one or both solutions in (3.44) or (3.45) depending on the location of the sound sources in the considered setting. Precisely, due to the properties of spherical Bessel and spherical Hankel functions, the following choices are widely accepted in the literature:

- the spherical Bessel function of the first kind, $j_l((\omega/c)r)$, is used to represent incoming acoustic fields, i.e. acoustic fields due to sources far from the origin and observed around the origin (cfr. the interior problem in Sec. 4.2.2);
- the spherical Hankel function of the first kind, $h_l((\omega/c)r)$ is used to represent outgoing acoustic fields, i.e. acoustic fields due to sources closed to the origin and observed in an external region (cfr. the exterior problem in Sec. 4.2.1).

The acoustic field $P(\mathbf{r},\omega) = R(r)\Phi(\phi)\Theta(\theta)$ is called *spherical wave*.

3.3.3 Cylindrical Waves

In cylindrical coordinates the Laplace operator is (3.9), hence the Helmholtz equation reads

$$\frac{\partial^2 P(\mathbf{r},\omega)}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial P(\mathbf{r},\omega)}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 P(\mathbf{r},\omega)}{\partial \phi^2} + \frac{\partial^2 P(\mathbf{r},\omega)}{\partial z^2} + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r},\omega) = 0. \tag{3.46}$$

Similarly to the spherical case, also in the cylindrical setting solutions to the Helmholtz equation are obtained by separation of variables. The function $P(\mathbf{r},\omega)$ is factored as

$$P(\mathbf{r},\omega) = R(\rho)\Phi(\phi)Z(z), \tag{3.47}$$

where $R(\rho)$, $\Phi(\phi)$ and $Z(z)$ denote the radial dependency, the angular dependency and the dependency on height z, respectively.

Upon substituting (3.47) into (3.46), taking constant terms out of the derivatives and dividing by $R(\rho)\Phi(\phi)Z(z)$ one obtains

$$\frac{1}{R(\rho)}\frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{1}{\rho R(\rho)}\frac{\partial R(\rho)}{\partial \rho} + \frac{1}{\rho^2\Phi(\phi)}\frac{\partial^2 \Phi(\phi)}{\partial \phi^2} + \frac{1}{Z(z)}\frac{\partial^2 Z(z)}{\partial z^2} + \left(\frac{\omega}{c}\right)^2 = 0. \tag{3.48}$$

Dependence on Height. By observing (3.48), one notices that the only term depending on the height z is the fourth, while the others can be regarded to as constant with respect to z. Thus, the height dependency is given by the solution to

$$\frac{1}{Z(z)}\frac{\partial^2 Z(z)}{\partial z^2} = -k_z^2, \tag{3.49}$$

where the constant $-k_z^2$ is set for convenience. The solution to (3.49) is given in Williams [1999, p. 117] as

$$Z(z) = Z_1 + e^{jk_z z} + Z_2 e^{-jk_z z}, \quad Z_1, Z_2 \in \mathbb{R}. \tag{3.50}$$

Dependence on Azimuth. Upon substituting (3.50) into (3.48) one obtains

$$\frac{1}{R(\rho)}\left(\frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial R(\rho)}{\partial \rho}\right) + \frac{1}{\rho^2\Phi(\phi)}\frac{\partial^2 \Phi(\phi)}{\partial \phi^2} = -\left(\frac{\omega}{c}\right)^2 + k_z^2 \triangleq -k_\rho^2, \tag{3.51}$$

where the quantity $k_\rho = \sqrt{(\omega/c)^2 - k_z^2}$ has been introduced. Upon multiplying (3.51) by ρ^2, it can be noticed that the only term depending on ϕ is the third, while the others can be regarded to as constants. The azimuthal dependency of the solution is thus given by

$$\frac{1}{\Phi(\phi)}\frac{\partial^2 \Phi(\phi)}{\partial \phi^2} = -m^2, \tag{3.52}$$

whose solution is [Williams, 1999, p. 117]

$$\Phi(\phi) = e^{jm\phi}, \quad m \in \mathbb{Z}. \tag{3.53}$$

Dependence on Radius. The radial dependency is governed by the equation

$$\frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial R(\rho)}{\partial \rho} + \left(k_\rho^2 - \frac{m^2}{\rho^2}\right) R(\rho) = 0, \tag{3.54}$$

which is recognized as the *Bessel equation* (also known as Riccati equation), whose solutions are Bessel's functions of the first and second kind [Olver, 2010, p. 217]

$$R(\rho) = R_1 J_m(k_\rho \rho) + R_2 Y_m(k_\rho \rho). \tag{3.55}$$

Alternatively, the solution to (3.54) can be expressed in terms of Hankel functions [Olver, 2010, p. 217] as in Williams [1999, p. 118], yielding

$$R(\rho) = R_3 H_m^{(1)}(k_\rho \rho) + R_4 H_m^{(2)}(k_\rho \rho), \tag{3.56}$$

where $H_m^{(1)}(\cdot)$ and $H_m^{(2)}(\cdot)$ denote the Hankel function of first kind [Olver, 2010, p. 217, Eq. 10.2.3-6] and second kind [Olver, 2010, p. 217, Eq. 10.2.6], respectively. The acoustic field $P(\mathbf{r}, \omega) = R(\rho)\Phi(\phi)Z(z)$ is called *cylindrical wave*.

Similarly to the spherical wave case, also in the case of cylindrical waves the Bessel functions of first kind are used to represent acoustic fields in a region internal with respect to a source distribution, while Hankel functions of the first kind are used to deal with the external scenario.

3.4 Solutions to the Inhomogeneous Wave Equation

The homogeneous wave equation describes acoustic fields in volumes free of sources. On the other hand, acoustic fields generated by sources inside the volume of interest are governed by the inhomogeneous wave equation (3.14) and the corresponding Helmholtz equation (3.15). In this section, mimicking what we have done for the homogeneous case, we assume time-harmonic dependency for the acoustic field; thus we seek for solutions to the inhomogeneous Helmholtz equation (3.15).

Arbitrary solutions to (3.15) are constructed starting from a basic solution, resulting by imposing a spatial impulse at location $\mathbf{r}'$ as the excitation term, i.e.

$$Q(\mathbf{r}, \omega) = \delta(\mathbf{r} - \mathbf{r}')\delta(\omega), \tag{3.57}$$

where $\delta(\cdot)$ denotes the Dirac delta function [Vetterli et al., 2014, pp. 316-317]. The sound field $G(\mathbf{r}|\mathbf{r}', \omega)$ measured at a point $\mathbf{r}$ due to a spatial impulse at position $\mathbf{r}'$ is solution to the inhomogeneous equation obtained upon substituting (3.57) into (3.15)

$$\nabla^2 G(\mathbf{r}|\mathbf{r}', \omega) + \left(\frac{\omega}{c}\right)^2 G(\mathbf{r}|\mathbf{r}', \omega) = -\delta(\mathbf{r} - \mathbf{r}'). \tag{3.58}$$

The function $G(\cdot)$ is called *Green function* and it has been formally defined in Morse and Feshbach [1953, p. 803, Sec. 7.2]. In a three-dimensional spatial domain and under free space conditions, the Green function is [Williams, 1999, p. 265, Eq. 8.41]

$$G(\mathbf{r}|\mathbf{r}', \omega) = \frac{e^{-j(\omega/c)\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r} - \mathbf{r}'\|}. \tag{3.59}$$

Equation (3.59) describes the acoustic field of a point source located at $\mathbf{r}'$ (cfr. Sec. 3.3.2), radiating energy in an omni-directional fashion (i.e. radiation only depends on the distance $\|\mathbf{r} - \mathbf{r}'\|$, not on angle).

In some situations that will appear during this thesis, it is useful to consider acoustic fields which are invariant over one spatial coordinate. Without loss of generality, we

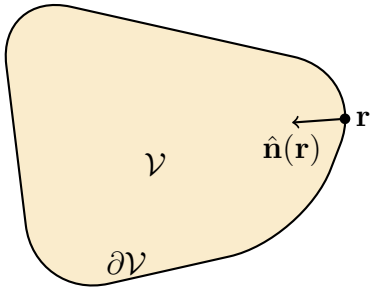


Figure 3.4: Geometry of a boundary value problem.

consider the case of acoustic fields invariant over the z coordinate: this class of acoustic fields is usually referred to as *height-invariant* acoustic fields and it is employed to derive simplifications of three-dimensional problems. In this context, the excitation function (3.57) can be written as

$$j\omega Q(\mathbf{r},\omega) = \delta(x-x')\delta(y-y'), \quad z \in \mathbb{R}. \tag{3.60}$$

This excitation function defines an infinite *line source* in the z direction located at (x', y'). The line source can be interpreted as the superposition of point sources along a straight line parallel to the z axis. The Green's function in this setting is obtained by integrating the Green's function of point sources along the z direction

$$G_1(\mathbf{r}|\mathbf{r}',\omega) = \int_{-\infty}^{\infty} G_0(\mathbf{r}|\mathbf{r}',\omega)\, dz', \quad \mathbf{r}' = [x', y', z']^T. \tag{3.61}$$

Upon substituting (3.59) into (3.61) one obtains [Williams, 1999, p. 266, Eq. 8.47]

$$G_1(\mathbf{r}|\mathbf{r}',\omega) = \int_{-\infty}^{\infty} \frac{e^{-j(\omega/c)\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|}\, dz' = \frac{j}{4} H_0^{(1)}\left(\frac{\omega}{c}\rho\right), \tag{3.62}$$

where the last equality follows from Gradshteyn and Ryzhik [2007, p. 915, Eq. 8.421.11] with $\rho = \sqrt{(x-x')^2 + (y-y')^2}$.

3.5 Boundary Conditions

In the previous sections the only scenario that has been considered is the free-field scenario, i.e. the Helmholtz equations (3.13) and (3.15) have been solved in free space. Although the free-space scenario is fundamental to build many of the techniques that are presented in this thesis, there are situations in which one has to consider also the presence of acoustic boundaries. Figure 3.4 illustrates this concept. The region $\mathcal{V}$ is the region of space where the acoustic field is governed by the Helmholtz equation, while on the boundary $\partial\mathcal{V}$ boundary conditions are specified. There exist several boundary conditions, according to the quantity that is constrained by the boundary, and to the type of constraint. In the following some of the most used boundary conditions are illustrated, each emerging from a specific acoustic behavior of the contour.

Dirichlet Boundary Condition. If the boundary constrains a value to the acoustic field itself, i.e. $p(\mathbf{r},\omega)$ is constrained for $\mathbf{r} \in \partial\mathcal{V}$, the boundary problem is called *Dirichlet boundary problem*. The simplest of these problems is the one in which the boundary

constrains the acoustic field to vanish. This condition is called *homogeneous Dirichlet boundary condition* [Ahrens, 2012, p. 49, Eq. 2.58]

$$P(\mathbf{r},\omega) = 0, \quad \forall \mathbf{r} \in \partial\mathcal{V} \tag{3.63}$$

and it describes a pressure-release boundary. Another problem is the one in which the acoustic field on the boundary is constrained to be equal to an arbitrary function: this condition is called *Inhomogeneous Dirichlet boundary condition* [Ahrens, 2012, p. 49, Eq. 2.59]

$$P(\mathbf{r},\omega) = F(\mathbf{r},\omega), \quad \forall \mathbf{r} \in \partial\mathcal{V}. \tag{3.64}$$

Neumann Boundary Condition. If the boundary dictates a constraint on the directional derivative of the acoustic field, the boundary problem is called *Neumann boundary problem*. If the boundary constrains the normal directional derivative of the acoustic field to vanish, the condition is called *Neumann boundary condition* [Ahrens, 2012, pp. 49-50, Eq. 2.60]

$$\langle \nabla P(\mathbf{r},\omega), \hat{\mathbf{n}}(\mathbf{r}) \rangle = 0, \quad \forall \mathbf{r} \in \partial\mathcal{V}, \tag{3.65}$$

where $\hat{\mathbf{n}}(\mathbf{r})$ denotes the unit vector normal to $\partial\mathcal{V}$ at a point $\mathbf{r} \in \partial\mathcal{V}$, as depicted in Fig. 3.4; this condition describes a sound hard boundary (like, e.g., a perfectly reflective wall). If the normal directional derivative is constrained to an arbitrary function on the boundary, the condition is called *inhomogeneous Neumann boundary condition* [Ahrens, 2012, p. 50, Eq. 2.63]

$$\langle \nabla P(\mathbf{r},\omega), \hat{\mathbf{n}}(\mathbf{r}) \rangle = F(\mathbf{r},\omega), \quad \forall \mathbf{r} \in \partial\mathcal{V}. \tag{3.66}$$

Robin Boundary Condition. In many practical situations the boundary is neither perfectly rigid nor pressure-release, thus neither of Dirichlet and Neumann boundary conditions accurately model the behavior of the boundary. In all situations where the boundary can be considered as an *absorbing boundary*, the *homogeneous Robin boundary condition* [Filippi et al., 1998, p. 44]

$$\langle \nabla P(\mathbf{r},\omega), \hat{\mathbf{n}}(\mathbf{r}) \rangle - j\frac{\omega}{c}\frac{1}{\gamma(\mathbf{r},\omega)} P(\mathbf{r},\omega) = 0 \tag{3.67}$$

can be adopted. This boundary condition states that the normal directional derivative of the acoustic field is proportional to the acoustic field itself. The quantity $\gamma(\mathbf{r},\omega)$ is called *specific normal impedance*.

The three boundary conditions seen so far can be unified into a single expression

$$\alpha \langle \nabla P(\mathbf{r},\omega), \hat{\mathbf{n}}(\mathbf{r}) \rangle + \beta P(\mathbf{r},\omega) = F(\mathbf{r},\omega), \quad \forall \mathbf{r} \in \partial\mathcal{V}, \tag{3.68}$$

with the following cases:

- $\alpha = 0$, $\beta = 1$ dictates a Dirichlet boundary condition;
- $\alpha = 1$, $\beta = 0$ dictates a Neumann boundary condition;
- $\alpha = 1$, $\beta \neq 0$ dictates a Robin boundary condition; and,
- $F(\mathbf{r},\omega) = 0$ dictates homogeneous boundary conditions.

In many situations (e.g. in the sound field reproduction problem), it is customary to consider a boundary valued problem with a homogeneous Helmholtz equation and an inhomogeneous boundary condition. This scenario models the practical case of acoustic energy being furnished into an area by a motion of the boundary.

Sommerfeld Boundary Condition. In the setting of external acoustic problems (cfr. Fig. 4.2) the *Sommerfeld condition* imposes a boundary condition when the boundary is at infinity. This condition states that no energy contribution to the acoustic field is originated at infinity [Ahrens, 2012, pp. 50-51, Eq. 2.64]

$$\lim_{r\to\infty} r\left(\frac{\partial P(\mathbf{r},\omega)}{\partial r} + j\frac{\omega}{c}P(\mathbf{r},\omega)\right) = 0, \tag{3.69}$$

where $r = \|\mathbf{r}\|$.

3.6 Integral Representation of Acoustic Fields

This section shows how the acoustic field in a confined spatial region can be represented as the radiation of sources located on its boundary. The topics reviewed in this section are of fundamental importance in this thesis, since they provide the basis for the acoustic field representations illustrated in Chap. 4 and then used in all the following chapters, both for acoustic field analysis and synthesis purposes.

3.6.1 Kirchoff-Helmholtz Integral Equation

Consider the volume $\mathcal{V}$ and its boundary $\partial\mathcal{V}$. Let $F(\mathbf{r},\omega)$ be the space-time excitation function of a time-harmonic source and $P(\mathbf{r},\omega)$ the corresponding acoustic field. The function $P(\mathbf{r},\omega)$ must be a solution to the boundary value problem

$$\nabla^2 P(\mathbf{r},\omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r},\omega) = F(\mathbf{r},\omega), \quad \mathbf{r}\in\mathcal{V} \tag{3.70}$$

$$\alpha\langle\nabla P(\mathbf{r},\omega), \hat{\mathbf{n}}(\mathbf{r})\rangle + \beta P(\mathbf{r},\omega) = 0, \quad \forall\mathbf{r}\in\partial\mathcal{V}. \tag{3.71}$$

The acoustic field $P(\mathbf{r},\omega)$ can be considered as the superposition of two contributions:

- the acoustic field $P_0(\mathbf{r},\omega)$ generated by the radiation of the source $F(\mathbf{r},\omega)$ in an unbounded domain, which is given by the Green function (3.59); and,
- the acoustic field $P_1(\mathbf{r},\omega)$ which is a solution to the homogeneous Helmholtz equation satisfying the boundary condition in 3.71; this contribution represents the scattered field, as pointed out in Filippi et al. [1998, p. 81].

The expression that represents a solution to (3.70) with the boundary condition (3.71) is usually referred to as *Kirchoff-Helmholtz integral equation* [Williams, 1999, p. 256], [Filippi et al., 1998, p. 84] and reads

$$a(\mathbf{r})P(\mathbf{r},\omega) = P_0(\mathbf{r},\omega) + \int_{\partial\mathcal{V}} \left(G(\mathbf{r}|\mathbf{r}',\omega)\langle\nabla P(\mathbf{r}',\omega), \hat{\mathbf{n}}(\mathbf{r}')\rangle - P(\mathbf{r}',\omega)\langle G(\mathbf{r}|\mathbf{r}',\omega), \hat{\mathbf{n}}(\mathbf{r}')\rangle\right) dA(\mathbf{r}'), \tag{3.72}$$

where $A(\mathbf{r}')$ is an infinitesimal element of the surface $\partial\mathcal{V}$ and $a(\mathbf{r})$ is the discrimination term [Williams, 1999, p. 256]

$$a(\mathbf{r}) = \begin{cases} 1, & \text{for } \mathbf{r}\in\mathcal{V} \\ 0.5, & \text{for } \mathbf{r}\in\partial\mathcal{V} \\ 0, & \text{for } \mathbf{r}\notin\mathcal{V}. \end{cases} \tag{3.73}$$

Equation (3.72) can be readily interpreted as follows. The acoustic field inside the region $\mathcal{V}$ is uniquely determined by three contributions:

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

It must be noticed that the two latter functions (the pressure $P(\mathbf{r}', \omega)$ and the directional derivative $\langle \nabla P(\mathbf{r}', \omega), \hat{\mathbf{n}}(\mathbf{r}') \rangle$) are not independent, since they are related by the boundary condition (3.71) [Filippi et al., 1998, p. 84].

The Kirchoff-Helmholtz integral is the basis for the derivation of Wave Field Synthesis, as derived in Berkhout et al. [1993], Spors et al. [2008]. The Kirchoff-Helmholtz integral is a powerful tool, since it allows one to derive physically-motivated acoustic field synthesis methodologies.

3.6.2 Single Layer Potential and Double Layer Potential

The Kirchoff-Helmholtz integral equation (3.72) contains two integral contributions

$$\varphi_1(\mathbf{r}, \omega) = \int_{\partial\mathcal{V}} G(\mathbf{r}|\mathbf{r}', \omega) \langle \nabla P(\mathbf{r}', \omega), \hat{\mathbf{n}}(\mathbf{r}') \rangle \, dA(\mathbf{r}'), \tag{3.74}$$

$$\varphi_2(\mathbf{r}, \omega) = -\int_{\partial\mathcal{V}} P(\mathbf{r}', \omega) \langle G(\mathbf{r}|\mathbf{r}', \omega), \hat{\mathbf{n}}(\mathbf{r}') \rangle \, dA(\mathbf{r}'). \tag{3.75}$$

The function $\varphi_1(\mathbf{r}, \omega)$ is to be interpreted as the acoustic field radiated by a distribution of point sources on $\partial\mathcal{V}$, each driven by the function $\langle P(\mathbf{r}', \omega), \hat{\mathbf{n}}(\mathbf{r}') \rangle$. The expression (3.74) is referred to as *single layer potential* [Colton and Kress, 1992, p. 38] or *simple source formulation* [Williams, 1999, p. 267], [Spors et al., 2008]. The function $\varphi_2(\mathbf{r}, \omega)$ can be interpreted as the acoustic field radiated by a distribution of dipole sources on $\partial\mathcal{V}$; for this reason it is usually called *double layer potential* [Filippi et al., 1998, p. 85].

Conditions for the existence and uniqueness of the solution to (3.72), (3.74) and (3.75) have been derived in Lalor [1968]; more recently they have been reviewed in Filippi et al. [1998, p. 87] and in Fazi and Nelson [2012].

3.7 Ray Acoustics

In some acoustic propagation problems, it may be impractical to solve exactly the wave equation, along with its boundary conditions, due to the enormous computational effort that would be required. Indeed, this is the case of acoustic propagation in realistic rooms, where all the room's details make the problem non tractable. Moreover, the exact solution may be difficult to interpret in terms of one's intuition about acoustic waves. This section reviews the theory supporting *geometrical acoustics*, derived starting from the high frequency approximation of the wave equation as proposed in Johnson and Dudgeon [1993, pp. 33-36].

Ray acoustics is the simplest approximation of wave acoustic theory. In its view, sound is described by rays traveling in accordance with a set of geometrical rules. Ray acoustics is concerned with the location and direction of acoustic rays. In the scenario considered in this thesis, where only homogeneous and isotropic propagation media are considered, acoustic rays point in the direction of flow of the acoustic energy. Thanks to this characteristic, acoustic rays are usually employed to study the directional propagation of sound. For the sake of an example, microphone arrays can be built with the purpose of estimating the acoustic energy reaching the array from a pre-determined set of directions.

This process is analogous to the *image formation* process in optics [Saleh and Teich, 1991, p. 3], where the collection of rays reaching an optical component is mapped onto a corresponding intensity image. Thank to this analogy, acoustic directional intensity maps are usually referred to as *acoustic images*. Another application based on acoustic rays involves their tracing in complex acoustic environments in order to predict the acoustic behavior of the environment itself [Vorländer, 2008, Chap. 11].

Ray acoustics is based on a set of postulates, similar to the postulates of *ray optics*. Let us first consider those postulates in the realm of optics, according to Saleh and Teich [1991, pp. 3-4].

1. Light travels in the form of rays. Rays are emitted by light sources and can be sensed by light receivers.

2. In an homogeneous medium (like air), in which the speed of light c is a function of temperature but can be assumed to be invariant in space. Therefore, the time taken by light to propagate over a distance d is $\tau = d/c$.

3. Light rays traveling between two points follow a path such that the time of travel is minimum (*Fermat's principle*), which corresponds also to the path of minimum length (*Hero's principle*). Therefore, light rays in air travel in straight lines.

4. Light radiance is invariant along a ray.

Upon introducing some assumptions, these postulates are readily applied to the acoustic realm. In the following, we introduce approximations to the homogeneous Helmholtz equation in order to derive a physical model for acoustic rays.

3.7.1 The Eikonal Equation

Ray trajectories can also by univocally identified by the surfaces $\psi(\mathbf{r}, \omega)$ to which they are normal. In this setting, ray trajectories can be constructed considering the direction of the gradient vector $\nabla\psi(\mathbf{r}, \omega)$ at each point $\mathbf{r}$. The function $\psi(\mathbf{r}, \omega)$ is usually referred to as *Eikonal* [Saleh and Teich, 1991, p. 25]. In the following, the equation governing Eikonals is derived starting from an high-frequency approximation of the Helmholtz equation.

Consider the Helmholtz equation (3.13), here rewritten for completeness

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = 0, \tag{3.76}$$

where the acoustic field $P(\mathbf{r}, \omega)$ is a complex function and thus can be written in terms of its magnitude and phase as

$$P(\mathbf{r}, \omega) = |P(\mathbf{r}, \omega)| e^{j\psi(\mathbf{r}, \omega)}. \tag{3.77}$$

Inserting (3.77) into (3.76) yields

$$\nabla^2 \left(|P| e^{j\psi}\right) + \left(\frac{\omega}{c}\right)^2 |P| e^{j\psi} = 0, \tag{3.78}$$

where, for brevity, the dependence of P and ψ on $\mathbf{r}$ and ω has been omitted. Expanding the Laplace operator for the product of the two functions $|P|$ and ψ and then applying its linearity property yields

$$\nabla \cdot \left[(\nabla |P|) e^{j\psi}\right] + \nabla \cdot \left[|P| \nabla e^{j\psi}\right] + \left(\frac{\omega}{c}\right)^2 |P| e^{j\psi} = 0, \tag{3.79}$$

and then

$$\left(\nabla^2|P|\right)e^{j\psi} + 2\langle\nabla|P|, \nabla e^{j\psi}\rangle + |P|\nabla^2 e^{j\psi} + \left(\frac{\omega}{c}\right)^2 |P|e^{j\psi} = 0. \tag{3.80}$$

Upon taking the gradient of $e^{j\psi}$ and then multiplying by $e^{-j\psi}$ one obtains [Johnson and Dudgeon, 1993, p. 34]

$$\nabla^2|P| + |P|\left\{\left(\frac{\omega}{c}\right)^2 - \langle\nabla\psi, \nabla\psi\rangle\right\} + j\left\{|P|\nabla^2\psi + 2\langle\nabla|P|, \nabla\psi\rangle\right\} = 0. \tag{3.81}$$

Assuming high frequency, i.e. $\omega \to \infty$, the first term in (3.81) becomes negligible. For the equation to be satisfied, both the second and the third terms must equal zero, since the second is purely real while the third is purely imaginary. Setting the second terms to zero yields the *Eikonal equation* as presented in Goodman [1996, p. 402]

$$\langle\nabla\psi, \nabla\psi\rangle = \left(\frac{\omega}{c}\right)^2, \tag{3.82}$$

being ψ the *Eikonal*. The physical interpretation of (3.82) is that it constrains acoustic rays to travel in the direction orthogonal to lines of constant phase. This observation will be expanded in the next paragraph.

It should be remarked that acoustic rays encode exactly the same information than the Eikonals, as these two quantities are intimately related by (3.82). Moreover, in the following paragraph we will show that acoustic rays encode the same information also with respect to plane waves: this motivates us in the use of acoustic rays to model acoustic fields that can be represented as a superposition of plane waves.

3.7.2 Rays in Acoustic Fields

Acoustic rays emerge as the lines orthogonal to the solution of (3.82). In free-space, the acoustic rays (normals to the Eikonal $\psi(\mathbf{r}, \omega)$) must be straight lines [Saleh and Teich, 1991, p. 26] and the Eikonal may be parallel planes or concentric sphere. The following paragraphs illustrate this concept.

Rays in Plane Wave Acoustic Fields. Consider a plane wave acoustic field propagating with wavenumber vector $\mathbf{k}$

$$P(\mathbf{r}, \omega) = A(\omega)e^{j\langle\mathbf{k},\mathbf{r}\rangle}, \tag{3.83}$$

where $A(\omega)$ is an amplitude function. Following (3.77), the plane wave acoustic field can be factorized as

$$P(\mathbf{r}, \omega) = |A(\omega)|e^{j(\langle\mathbf{k},\mathbf{r}\rangle+\phi)}, \tag{3.84}$$

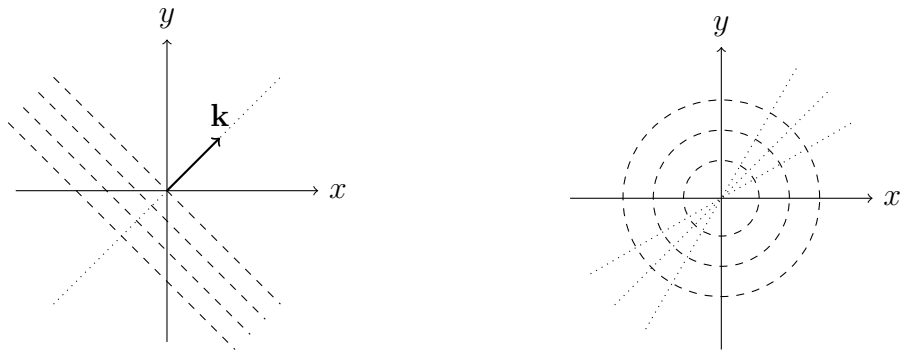
being $A(\omega) = |A(\omega)| + e^{j\phi}$. The Eikonal is

$$\psi(\mathbf{r}, \omega) = \langle\mathbf{k}, \mathbf{r}\rangle + \phi. \tag{3.85}$$

Upon substituting (3.85) into (3.82), it can be immediately seen that a plane wave is a solution to the Eikonal equation as long as the wavenumber vector $\mathbf{k}$ satisfies the dispersion relation $\mathbf{k} = (\omega/c)^2$. The trajectory of acoustic rays is determined by the gradient of $\psi(\mathbf{r}, \omega)$, which reads

$$\nabla\psi(\mathbf{r}, \omega) = \mathbf{k}. \tag{3.86}$$

Equation (3.86) establishes the strong analogy between the wavenumber vector and the trajectory of acoustic rays. This concept is illustrated in Fig. 3.5a.



(a) Rays in plane wave acoustic field. **(b)** Rays in spherical wave acoustic field.

Figure 3.5: Analogy between the wavenumber vector and acoustic rays in a plane wave acoustic field (Fig. 3.5a); the dashed lines represent the Eikonals, while the dotted line represents the trajectory of the acoustic ray. Rays in a spherical wave acoustic field (Fig. 3.5b) ; the dashed circles represent the Eikonals, while the dotted line represents the trajectories of the acoustic rays.

Rays in Spherical Wave Acoustic Fields. Consider a spherical wave acoustic field generated by a point source in the origin of the coordinate system

$$P(\mathbf{r},\omega) = \frac{e^{j\frac{\omega}{c}\|\mathbf{r}\|}}{4\pi\|\mathbf{r}\|}. \tag{3.87}$$

The Eikonal is

$$\psi(\mathbf{r},\omega) = \frac{\omega}{c}\|\mathbf{r}\|, \tag{3.88}$$

whose gradient is

$$\nabla\psi(\mathbf{r},\omega) = \frac{\omega}{c}\left(\frac{x}{\|\mathbf{r}\|}\hat{\mathbf{x}} + \frac{y}{\|\mathbf{r}\|}\hat{\mathbf{y}} + \frac{z}{\|\mathbf{r}\|}\hat{\mathbf{z}}\right). \tag{3.89}$$

Upon substituting (3.88) into (3.82) it is possible to verify that spherical wavefronts are solutions to the Eikonal equation. In this scenario, trajectories of acoustic rays belong to the bundle of lines crossing at the origin. Figure 3.5b illustrates this setting. In the more general scenario of a point source placed at $\mathbf{r}'$, acoustic rays belong to the bundle of lines crossing at $\mathbf{r}'$.

3.8 Beam Acoustics

This section introduces a physical model for acoustic fields in which sound is spatially confined and it propagates without angular spread. However, elementary solutions to the Helmholtz equation, like plane waves and spherical waves, seems to preclude such modelization. Rays in a plane wave acoustic field are aligned with the direction of propagation: in this setting there is no angular spread, but acoustic energy extends over the entire space. On the other hand, in a spherical wave acoustic fields rays originate from a single point, but they propagate diverging in all directions.

Although these two simple considerations (along with the wave acoustic theory reviewed in the preceding sections) seem to preclude the possibility of adopting a local model, in the following a physical model based on spatially localized wave objects is presented. The instance of such wave objects that is considered in this work is called *Gaussian*

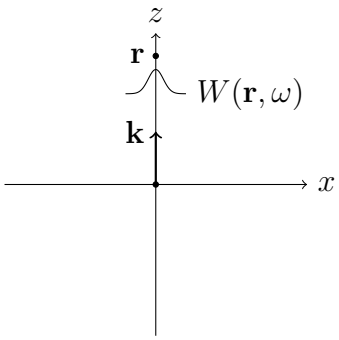


Figure 3.6: Ray approximation and paraxial approximation.

beam, for which the acoustic energy is concentrated principally withing a small region about the beam axis, while the energy distribution in any transverse plane is a Gaussian function centered about the beam axis itself. In other words, Gaussian beams have limited transverse extension. It should be remarked that such modelization is different from the concept of a beam used in geometrical acoustics. Indeed, in that domain, a beam is conceived as a bundle of rays originating from a point source; on the other hand, in this context, a beam is conceived as originating from a region of finite extent, not from an infinitesimal point source. In the following of this thesis it will be made clear how such difference plays an fundamental role.

This section, following the work in Saleh and Teich [1991, Chap. 3] and Goldsmith [1998, Chap. 2], provides a derivation for Gaussian beams starting from an approximated form of the Helmholtz equation; then, the physical properties of acoustic Gaussian beams are discussed.

3.8.1 The Paraxial Approximation

The paraxial approximation is an approximation to the Helmholtz equation that is valid for small angles from the ray axis. It supports an increase of acoustic modeling accuracy with respect to the ray-based model. Consider the illustrative scenario depicted in Fig. 3.6, where an acoustic source is placed in the origin and an acoustic receiver is on the z-axis at $\mathbf{r} = [0, 0, z]^T$. Within the acoustic ray approximation, the acoustic field received at $\mathbf{r}$ due to a source in $\mathbf{r}' = [0, 0, 0]^T$ is the field of a plane wave with wavenumber vector $\mathbf{k} = (\omega/c)\hat{\mathbf{z}}$

$$P(\mathbf{r}, \omega) \approx e^{j\frac{\omega}{c}z}, \tag{3.90}$$

meaning that only a single plane wave in the direction of $\mathbf{r}$ contributes to the field. A more refined approximation can be built associating to the propagation in the direction of $\mathbf{r}$ a complex profile $W(\mathbf{r}, \omega)$, such that the field received at $\mathbf{r}$ is

$$P(\mathbf{r}, \omega) = W(\mathbf{r}, \omega)e^{j\frac{\omega}{c}z}, \tag{3.91}$$

as it is illustrated in Fig. 3.6. The ray approximation is a specific instance of paraxial approximation with $W(\mathbf{r}, \omega) = \delta(\mathbf{r})$. Wave objects of the form (3.91) are called *paraxial waves* [Saleh and Teich, 1991, p. 50]. The variation of $W(\mathbf{r}, \omega)$ with position must be slow within the distance of a wavelength $\lambda = 2\pi/(\omega/c)$; in this case, the wave object maintains its local plane-wave nature [Saleh and Teich, 1991, p. 50]. The last assumption determines that within a wavelength distance $\delta z = \lambda$, the variation δW is much smaller than W itself.

Since $\delta W = (\partial W/\partial z)\delta z = (\partial W/\partial z)\lambda$, it follows [Saleh and Teich, 1991, p. 50, Eq. 2.2-20]

$$\frac{\partial W(\mathbf{r},\omega)}{\partial z} \ll \frac{\omega}{c} W(\mathbf{r},\omega), \tag{3.92}$$

and also the derivative $\partial W/\partial z$ varies slowly within a wavelength distance, so that

$$\frac{\partial^2 W(\mathbf{r},\omega)}{\partial z^2} \ll \frac{\omega}{c}\frac{\partial W(\mathbf{r},\omega)}{\partial z} \ll \left(\frac{\omega}{c}\right)^2 W(\mathbf{r},\omega). \tag{3.93}$$

Upon substituting (3.91) into (3.13) one obtains

$$\nabla^2 \left(W(\mathbf{r},\omega)e^{j\frac{\omega}{c}z}\right) + \left(\frac{\omega}{c}\right)^2 W(\mathbf{r},\omega)e^{j\frac{\omega}{c}z} = 0. \tag{3.94}$$

The expansion of the Laplace operator and the elimination of opposite terms yields

$$(\nabla^2 W(\mathbf{r},\omega))e^{j\frac{\omega}{c}z} + 2\langle \nabla W(\mathbf{r},\omega), \nabla e^{j\frac{\omega}{c}z}\rangle = 0 \tag{3.95}$$

The inner product in (3.95) can be simplified by computing

$$\nabla e^{j\frac{\omega}{c}z} = j\frac{\omega}{c}e^{j\frac{\omega}{c}z}\hat{\mathbf{z}}, \tag{3.96}$$

thus yielding

$$\langle \nabla W(\mathbf{r},\omega), \nabla e^{j\frac{\omega}{c}z}\rangle = j\frac{\omega}{c}\frac{\partial W(\mathbf{r},\omega)}{\partial z}e^{j\frac{\omega}{c}z}. \tag{3.97}$$

After substituting (3.97) into (3.95) and after multiplying by $e^{-j(\omega/c)z}$, (3.95) takes the form of the *reduced Helmholtz equation* as presented in Goldsmith [1998, p. 10, Eq. 2.4]

$$\nabla_T^2 W(\mathbf{r},\omega) + \frac{\partial^2 W(\mathbf{r},\omega)}{\partial z^2} + j2\frac{\omega}{c}\frac{\partial W(\mathbf{r},\omega)}{\partial z} = 0, \tag{3.98}$$

where the notation ∇_T^2 denotes the *transversal Laplace operator*

$$\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \tag{3.99}$$

In the sense of (3.93) one can neglect the factor $\partial^2 W/\partial z^2$ in comparison with $\partial W/\partial z$, thus obtaining

$$\nabla_T^2 W(\mathbf{r},\omega) + j2\frac{\omega}{c}\frac{\partial W(\mathbf{r},\omega)}{\partial z} = 0, \tag{3.100}$$

which is known as the *paraxial Helmholtz equation* [Saleh and Teich, 1991, p. 51, Eq. 2.2-22], [Goldsmith, 1998, p. 10, Eq. 2.5].

3.8.2 Gaussian Beams

Gaussian Beam Solution in Cylindrical Coordinates. In order to introduce the Gaussian beam solution to the paraxial Helmholtz equation in a progressive fashion, first the axial symmetric case (about the z axis) is considered. In particular, the cylindrical reference frame is adopted, being $\rho = \sqrt{x^2 + y^2}$ the radial distance from the z axis and ϕ being the angular coordinate, as depicted in Fig. 3.1c. In this coordinate system, the transversal Laplace operator becomes

$$\nabla_T^2 = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho} + \frac{1}{\rho}\frac{\partial^2}{\partial \phi^2}. \tag{3.101}$$

Hence, the paraxial Helmholtz equation (3.100) is written as

$$\frac{\partial^2 W(\mathbf{r},\omega)}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial W(\mathbf{r},\omega)}{\partial \rho} + \frac{1}{\rho}\frac{\partial^2 W(\mathbf{r},\omega)}{\partial \phi^2} + j2\frac{\omega}{c}\frac{\partial W(\mathbf{r},\omega)}{\partial z} = 0, \tag{3.102}$$

being $W(\mathbf{r},\omega) = W(\rho,\phi,z,\omega)$. Assuming axial symmetry (about z) makes $W(\rho,\phi,z,\omega)$ independent on ϕ, hence the second-order derivative of W with respect to ϕ in (3.102) equals zero. Thus, the *axially symmetric paraxial Helmholtz equation* as presented in Goldsmith [1998, p. 11, Eq. 2.7] reads

$$\frac{\partial^2 W(\rho,z,\omega)}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial W(\rho,z,\omega)}{\partial \rho} + j2\frac{\omega}{c}\frac{\partial W(\rho,z,\omega)}{\partial z} = 0. \tag{3.103}$$

Following Kogelnik and Li [1966, p. 1315, Eq. 12] in the analogy between (3.103) and the time dependent Schrödinger equation [Schrödinger, 1926], a trial solution to (3.103) is assumed in the form

$$W(\rho,z,\omega) = e^{j\left(\frac{\omega}{c}\frac{\rho^2}{2q(z)} - \psi(z)\right)}, \tag{3.104}$$

being $\psi(z)$ and $q(z)$ complex function to be determined. In order to substitute (3.104) into (3.102), first the single derivatives in (3.102) are computed, yielding

$$\frac{\partial W(\rho,z,\omega)}{\partial \rho} = j\frac{\omega}{c}\frac{\rho}{q(z)}W(\rho,z,\omega), \tag{3.105}$$

$$\frac{\partial^2 W(\rho,z,\omega)}{\partial \rho^2} = \left[j\frac{\omega}{c}\frac{1}{q(z)} - \left(\frac{\omega}{c}\right)^2\frac{\rho^2}{q^2(z)}\right]W(\rho,z,\omega), \tag{3.106}$$

$$\frac{\partial W(\rho,z,\omega)}{\partial z} = -\left[j\frac{\omega}{c}\frac{\rho^2}{2q^2(z)}\frac{dq(z)}{dz} + j\frac{d\psi(z)}{dz}\right]W(\rho,z,\omega). \tag{3.107}$$

Upon substituting (3.105)-(3.107) into (3.102) one obtains

$$-\left(\frac{\omega}{c}\right)^2\frac{\rho^2}{q^2(z)}\left[1 - \frac{dq(z)}{dz}\right]W(\rho,z,\omega) + 2\frac{\omega}{c}\left[\frac{d\psi(z)}{dz} + \frac{j}{q(z)}\right]W(\rho,z,\omega) = 0. \tag{3.108}$$

From (3.108) one can infer that the functions $q(z)$ and $\psi(z)$ must satisfy the differential equations

$$\frac{dq(z)}{dz} = 1, \tag{3.109}$$

$$\frac{d\psi(z)}{dz} = -\frac{j}{q(z)}. \tag{3.110}$$

The solution to (3.109) is of the form

$$q(z) = z + q(0), \tag{3.111}$$

where $q(0)$ denotes the initial value of $q(z)$ for $z = 0$. It is evident that $q(0)$ must be a measure of length, so that its dimension is the same as z; however, $q(0)$ can not be a real value. To demonstrate this statement, consider *ad absurdum* $q(0) \in \mathbb{R}$; then

$$W(\rho,z,\omega) = e^{j\frac{\omega}{c}\frac{\rho^2}{2q^2(z)}}e^{-j\psi(z)}. \tag{3.112}$$

Since $|e^{j\frac{\omega}{c}\frac{\rho^2}{2q^2(z)}}| = 1$ and since $\psi(z)$ is independent on ρ, one concludes that the phase of $W(\rho,z,\omega)$ is highly oscillatory with ρ, but its amplitude remains constant. This situation

does not describe a beam, which is conceived to have most of its energy concentrated near the beam axis (i.e. $\rho \approx 0$) and gradually decaying as ρ increases. Hence, the case of having $q(0) \in \mathbb{R}$ must be discarded. On the other hand, if $q(0)$ is purely imaginary, i.e. $q(0) = jz_0$ one can write

$$q(z) = z + jz_0. \tag{3.113}$$

Since $q(z)$ is present in (3.104) at the denominator of the exponential function, it is worth expressing it as

$$\frac{1}{q(z)} = \frac{1}{z + jz_0} = \frac{z}{z^2 + z_0^2} - j\frac{z_0}{z^2 + z_0^2}; \tag{3.114}$$

thus $W(\rho, z, \omega)$ becomes

$$W(\rho, z, \omega) = e^{j\frac{\omega}{c}\frac{z\rho^2}{2(z^2+z_0^2)}} e^{\frac{\omega}{c}\frac{z_0\rho^2}{2(z^2+z_0^2))}} e^{-j\psi(z)}. \tag{3.115}$$

Consider now the solution to (3.110). The substitution of (3.113) into (3.110) yields

$$\frac{d\psi(z)}{dz} = -\frac{j}{z + jz_0}, \tag{3.116}$$

whose solution is given by

$$\begin{aligned} \psi(z) &= -j\int_0^z \frac{1}{z' + jz_0}\, dz' = \log\left(z' + jz_0\right)\Big|_0^z \\ &= -j\log\left(1 - j\frac{z}{z_0}\right). \end{aligned} \tag{3.117}$$

Hence, the term $e^{-j\psi(z)}$ in (3.115) becomes

$$e^{-j\psi(z)} = e^{-\log(1-jz/z_0)} = \frac{1}{1 - j\frac{z}{z_0}}. \tag{3.118}$$

Upon writing the complex term $1 - j(z/z_0)$ in polar form, i.e.

$$1 - j\frac{z}{z_0} = \sqrt{1 + \left(\frac{z}{z_0}\right)^2}\, e^{-j\arctan\left(\frac{z}{z_0}\right)}, \tag{3.119}$$

(3.118) becomes

$$e^{-j\psi(z)} = \frac{e^{j\arctan(z/z_0)}}{\sqrt{1 + (z/z_0)^2}}. \tag{3.120}$$

Finally, upon substituting (3.120) into (3.115) and, in turn, substituting the result into (3.91) one obtains the expression for the acoustic field of an axis symmetric Gaussian beam about the z axis

$$P(\mathbf{r}, \omega) = e^{j\frac{\omega}{c}\frac{z\rho^2}{2(z^2+z_0^2)}} e^{\frac{\omega}{c}\frac{z_0\rho^2}{2(z^2+z_0^2)}} \frac{e^{j\arctan(z/z_0)}}{\sqrt{1 + (z/z_0)^2}} e^{j\frac{\omega}{c}z}. \tag{3.121}$$

In order for (3.121) to represent a physically meaningful wave object, it must be ensured that the real exponential term is decaying; this implies $z_0 \leq 0$; for convenience, a new variable $z_c = -z_0$ is introduced and after this substitution, (3.122) becomes

$$P(\mathbf{r}, \omega) = e^{j\frac{\omega}{c}\frac{z\rho^2}{2(z^2+z_c^2)}} e^{-\frac{\omega}{c}\frac{z_c\rho^2}{2(z^2+z_c^2)}} \frac{e^{-j\arctan(z/z_c)}}{\sqrt{1 + (z/z_c)^2}} e^{j\frac{\omega}{c}z}. \tag{3.122}$$

Beam Parameters. In order to provide an illustrative view of beam wave objects, some fundamental beam parameters are derived. First, consider the acoustic field of a beam on the $z = 0$ plane, i.e.

$$P(\rho, 0, \omega) = e^{-\frac{\omega}{c}\frac{z_c\rho^2}{2z_c^2}} = e^{-\frac{\omega}{c}\frac{\rho^2}{2z_c}}. \tag{3.123}$$

Equation (3.123) has the form of the Gaussian function

$$e^{-\frac{\rho^2}{2\sigma_0^2}}, \tag{3.124}$$

where the parameter σ_0^2 is related to the full width at half maximum FWHM[3] of the Gaussian function by

$$\text{FWHM} = 2\sqrt{2\log(2)}\sigma_0. \tag{3.125}$$

The parameter σ_0 is interpreted as the radius of the beam on the plane $z = 0$ and by comparison of (3.123) and (3.124) it is given by

$$\sigma_0 = \sqrt{\frac{z_c}{(\omega/c)}}. \tag{3.126}$$

Another parameter is the *radius of curvature* of the wavefronts in the beam wave object. This parameter is defined as

$$R(z) = z + \frac{z_c^2}{z}, \tag{3.127}$$

which in view of (3.126) can be rewritten as

$$R(z) = z + \frac{1}{z}\left(\frac{\omega}{c}\right)^2\left(\frac{\sigma_0^2}{2}\right)^2. \tag{3.128}$$

While the quantity σ_0 is interpreted as the radius of the beam on the plane $z = 0$, it is useful to obtain an analytic expression for the beam radius at arbitrary distances $z \neq 0$. This quantity is obtain in a way analogous to σ_0 removing the constraint of $z = 0$. Consider the real exponential term in (3.122); this term in in the form (3.124) where the parameter $\sigma(z)$ is now a function of z and can be obtained by equating the exponents in the two above mentioned equations

$$-\frac{\omega}{c}\frac{z_c\rho^2}{2(z^2 + z_c^2)} = \frac{\rho^2}{\sigma^2(z)}, \tag{3.129}$$

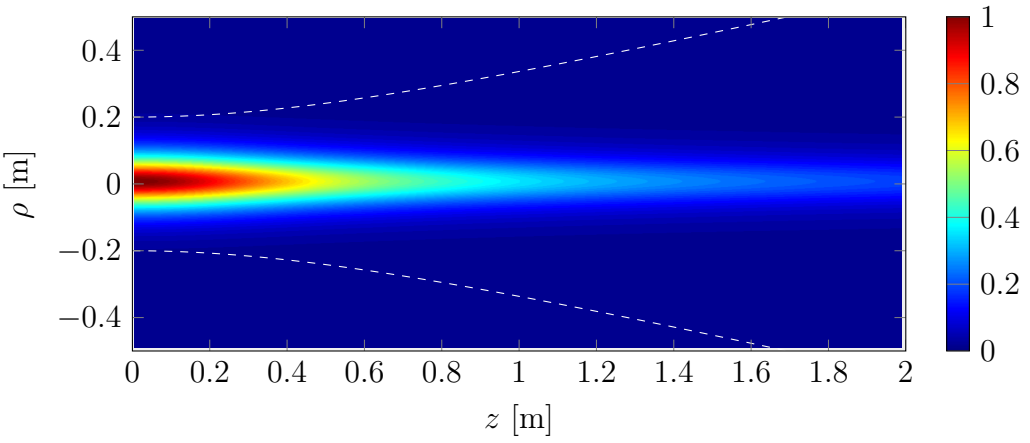
yielding

$$\sigma^2(z) = \frac{2}{(\omega/c)}\left(z_c + \frac{z^2}{z_c}\right). \tag{3.130}$$

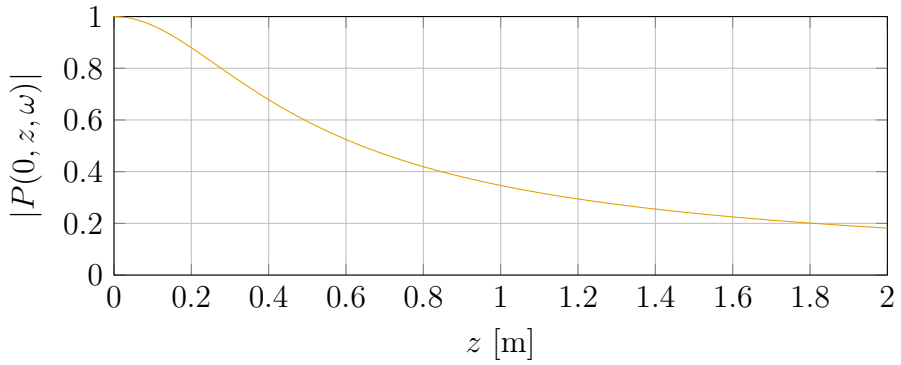
Upon substituting in (3.130) the expression for z_c that can be obtained from (3.126), i.e.

$$z_c = \frac{\omega}{c}\sigma_0^2, \tag{3.131}$$

[3]The full width at half maximum is defined here as the difference between the extreme values of ρ at which the Gaussian function is equal to half of its maximum. In the signal processing context, this measure is usually regarded to as the *bandwidth* of a function, being the extreme values the points at $-3\,\text{dB}$ of attenuation.



(a) Beam field.



(b) On-axis amplitude drop.

Figure 3.7: Transverse spreading and drop of the on-axis amplitude for axis symmetric beam propagation. White dashed lines in Fig. 3.7a denote the beam radius $\sigma(z)$. The temporal frequency is set to 1 kHz, the speed of sound is 340 m s^{-1} and the beam waist radius is set to $\sigma_0 = 0.05$ m.

and taking the square root, one obtains the *beam radius* as

$$\sigma(z) = \sigma_0 \sqrt{1 + \left(\frac{2z}{(\omega/c)\sigma_0^2} \right)^2}. \tag{3.132}$$

It can be directly observed from (3.132) that the beam radius is minimum at $z = 0$; hence the quantity σ_0 is usually referred to as *beam waist radius*. It can also be observed that at the beam waist (i.e. at $z = 0$) the radius of curvature of the wavefronts (3.128) is infinite; this means that at the beam waist the beam wavefronts are planar.

Adopting the beam parameters introduced in these paragraphs, (3.122) can be rewritten as

$$P(\mathbf{r}, \omega) = e^{j\frac{\omega}{c}\frac{\rho^2}{R(z)}} e^{-\frac{\omega}{c}\frac{\rho^2}{\sigma^2(z)}} \frac{e^{-j\arctan(2z/((\omega/c)\sigma_0^2))}}{\sqrt{1 + (2z/((\omega/c)\sigma_0^2))^2}} e^{j\frac{\omega}{c}z}. \tag{3.133}$$

For illustrative purposes, Fig. 3.7 shows the beam field and the drop of on-axis amplitude for a Gaussian beam of the form (3.133). The temporal frequency is set to 1 kHz, the speed of sound is 340 m s^{-1} and the beam waist radius is set to $\sigma_0 = 0.05$ m.

Rayleigh Range In the literature on Gaussian beams, it is common to write the beam parameters in term of another parameter, called *Rayleigh range* [Goldsmith, 1998, p. 22]. The Rayleigh range is defined in Goldsmith [1998, p. 22, Eq. 2.41] as

$$z_{\mathrm{R}} = \frac{\sigma_0^2}{2(\omega/c)}. \tag{3.134}$$

Adopting the definition (3.134), the beam parameters can be written as

$$R(z) = z + \frac{z_{\mathrm{R}}^2}{z}, \tag{3.135}$$

$$\sigma(z) = \sigma_0 \sqrt{1 + \left(\frac{z}{z_{\mathrm{R}}}\right)^2}, \tag{3.136}$$

$$e^{-j \arctan(2z/((\omega/c)\sigma_0^2))} = e^{-j \arctan(z/z_{\mathrm{R}})}. \tag{3.137}$$

At the beam waist, the beam radius $\sigma(z)$ attains its minimum value σ_0; here the radius of curvature is infinite, being the beam wavefronts planar at the beam waist. The phase term in (3.137) can be interpreted as the on axis phase shift between a Gaussian beam and a plane wave [Goldsmith, 1998, p. 23]; it is easy to infer from (3.137) that the phase shift is zero at the beam waist.

The beam radius increases monotonically with the distance z. It can be verified that for $z < z_{\mathrm{R}}$ the beam radius remains essentially constant, being $\sigma(z) \leq \sqrt{2}\sigma_0$ in that region. Due to this consideration, the Rayleigh range is usually regarded to as the propagation distance z within which the Gaussian beam propagates without significant spreading; beams in this range are called in the literature *collimated beams*.

Gaussian Beam Solution in 2-D Cartesian Coordinates. Before introducing the more general form of the Gaussian beam is three-dimension, this paragraph considers the simplified case of a beam field that is invariant on the y spatial coordinate. The paraxial Helmholtz equation (3.100) becomes

$$\frac{\partial^2 W(x,z,\omega)}{\partial x^2} + j2\frac{\omega}{c}\frac{\partial W(x,z,\omega)}{\partial z} = 0. \tag{3.138}$$

Assuming a solution of the form

$$W(x,z,\omega) = A_x(z) e^{j\frac{\omega}{c}\frac{x^2}{2q_x(z)}} \tag{3.139}$$

leads to the conditions

$$\frac{\partial q_x(z)}{\partial z} = 1, \tag{3.140}$$

$$\frac{\partial A_x(z)}{\partial z} = -\frac{1}{2}\frac{A_x(z)}{q_x(z)}. \tag{3.141}$$

The first equation is identical to (3.109), thus the same solution (3.113) is employed. This leads to the same definitions for the beam parameters as in the cylindrically-symmetric case. On the other hand, the solution for $A_x(z)$ can be assumed of the form [Goldsmith, 1998, p. 16]

$$\frac{A_x(z)}{A_x(0)} = \sqrt{\frac{q_x(0)}{q_x(z)}}, \tag{3.142}$$

showing that the real part of the solution, in the two-dimensional case, depends on the square root of z, while the phase shift is the same as in the previous case. Finally, the acoustic field due to a two-dimensional beam is [Goldsmith, 1998, p. 16, Eq. 2.30]

$$P(x,z,\omega) = \sqrt[4]{\frac{2}{\pi\sigma_x^2(z)}} e^{j\frac{\omega}{c}\frac{x^2}{R_x(z)}} e^{-\frac{\omega}{c}\frac{x^2}{\sigma_x^2(z)}} e^{-j\arctan(2z/((\omega/c)\sigma_{0,x}^2))} e^{j\frac{\omega}{c}z}. \tag{3.143}$$

Gaussian Beam Solution in 3-D Cartesian Coordinates. The full 3-D paraxial Helmholtz equation (3.100) can be solved by the method of separation of variables, assuming a solution given by the product of 2-D solutions of the form (3.139), i.e.

$$W(x,y,z,\omega) = A_x(z)A_y(z) e^{j\frac{\omega}{c}\frac{x^2}{2q_x(z)}} e^{j\frac{\omega}{c}\frac{y^2}{2q_y(z)}}, \tag{3.144}$$

yielding to the conditions

$$\frac{\partial q_x(z)}{\partial z} = 1, \quad \frac{\partial q_y(z)}{\partial z} = 1, \tag{3.145}$$

$$\frac{\partial A_x(z)}{\partial z} = -\frac{1}{2}\frac{A_x(z)}{q_x(z)}, \quad \frac{\partial A_y(z)}{\partial z} = -\frac{1}{2}\frac{A_y(z)}{q_y(z)}. \tag{3.146}$$

Being these conditions of the same form of those seen in the previous paragraph, the overall acoustic field due to a 3-D Gaussian beam is given by the product of its x and y independent components

$$\begin{aligned} P(x,y,z,\omega) = &\sqrt{\frac{2}{\pi\sigma_x(z)\sigma_y(z)}} e^{j\frac{\omega}{c}\frac{x^2}{R_x(z)}} e^{j\frac{\omega}{c}\frac{y^2}{R_y(z)}} e^{-\frac{\omega}{c}\frac{x^2}{\sigma_x^2(z)}} e^{-\frac{\omega}{c}\frac{y^2}{\sigma_y^2(z)}} \\ &e^{-j\arctan(2z/((\omega/c)\sigma_{0,x}^2))} e^{-j\arctan(2z/((\omega/c)\sigma_{0,y}^2))} e^{j\frac{\omega}{c}z}, \end{aligned} \tag{3.147}$$

being the beam radii

$$\sigma_x(z) = \sigma_{0,x}\sqrt{1+\left(\frac{2z}{(\omega/c)\sigma_{0,x}^2}\right)^2}, \tag{3.148}$$

$$\sigma_y(z) = \sigma_{0,y}\sqrt{1+\left(\frac{2z}{(\omega/c)\sigma_{0,y}^2}\right)^2} \tag{3.149}$$

and the wavefront curvatures

$$R_x(z) = z + \frac{1}{z}\left(\frac{\omega}{c}\right)^2\left(\frac{\sigma_{0,x}^2}{2}\right)^2, \tag{3.150}$$

$$R_y(z) = z + \frac{1}{z}\left(\frac{\omega}{c}\right)^2\left(\frac{\sigma_{0,y}^2}{2}\right)^2. \tag{3.151}$$

3.9 Summary

- Fundamental equations governing acoustic fields are introduced, both in the time-domain (referred to as homogeneous (3.10) and inhomogeneous (3.14) wave equations) and in the temporal-frequency domain (referred to as homogeneous (3.13) and inhomogeneous (3.15) Helmholtz equations).

- Plane waves (3.16), spherical waves (3.27) and cylindrical waves (3.50) are introduced as fundamental solutions to the homogeneous Helmholtz equation in Cartesian, spherical and cylindrical coordinate systems, respectively.
- The formalism of Green's functions is introduced as the solution to the inhomogeneous Helmholtz equation.
- The Kirchoff-Helmholtz integral equation (3.72) is introduced as a mathematical tool to determine the acoustic field inside a given a distribution of acoustic monopoles and dipoles on the boundary. A simplification to the Kirchoff-Helmholtz integral equation is provided by the single layer potential (3.74) formulation, which employs only acoustic monopoles.
- Ray acoustics is introduced as a powerful approximation to wave acoustic theory, valid at high temporal frequencies. Rays are formally introduced as the solution to the Eikonal equation (3.82), which is derived as an high-frequency approximation to the Helmholtz equation.
- Beam acoustic is introduced as a more accurate approximation with respect to ray acoustics. In particular, Gaussian beams are introduced as the solution to the paraxial Helmholtz equation (3.100), derived as a small-angle approximation to the Helmholtz equation.

This page intentionally left blank.

Chapter 4

Acoustic Field Representations

This chapter serves as a bridge between the physical acoustics perspective reviewed in Chap. 3 and the discipline of signal processing. In particular, this chapter presents representations for acoustic fields that are direct generalizations of the signal representations reviewed in Chap. 2. In this view, the attention is focused to *non-parametric* acoustic field representations, since these require no a-priori knowledge of the sound scene. In this scenario the sound scene is completely described by the concept of the acoustic field, already defined in Chap. 3 as a real-valued scalar function of spatial position and time, i.e.

$$p(\mathbf{r}, \omega), \quad \mathbf{r} \in \mathbb{R}^3, \quad t \in \mathbb{R}. \tag{4.1}$$

A naïve acoustic field representation would be to encode the value of the pressure at any instant of time in any position of space. Of course, this representation is impractical, since it requires the encoding of a prohibitive amount of information. The objective behind any acoustic field representation is to encode all the information contained in the acoustic field in a more manageable format.

The following steps are applied in order to present each of the acoustic field representations reviewed in this chapter.

1. A basis solution to the Helmholtz equation is chosen and a set of independent variables that determine the basis solution is identified.
2. General solutions to the Helmholtz equation are identified as a weighed superposition of a set of basis solutions.
3. Weighing coefficients are found, which depend on the set of independent variables.
4. The acoustic field is approximated through a weighed sum of basis solutions.

For simplicity, all the representations considered here target time-harmonic signals; for this reason, in the following just solutions to the Helmholtz equation are considered. In later chapters, for each of the specific applications considered, it will be explained how the representation is adapted to accommodate wide-band signals.

4.1 Plane Wave Representations

Upon choosing plane waves as basis solutions, one obtains that an arbitrary acoustic field satisfying the homogeneous Helmholtz equation can be represented as an integral expansion of plane waves propagating in all directions. The importance of such representation arise from several factors, pointed out in Devaney and Sherman [1973, p. 765]:

- plane waves have a simple functional form, the only independent variables that completely characterize a plane wave are direction of propagation and temporal frequency, thus their physical interpretation is straightforward;

- under certain conditions, the plane wave expansion can be inverted: indeed, plane waves amplitudes can be computed starting from the knowledge of the acoustic field; thus, the plane wave representation constitute an useful model in both analysis and synthesis problems.

Two different plane wave representations are discussed in this section. The first has been historically attributed to Whittaker [1903], who showed that an acoustic field satisfying the homogeneous Helmholtz equation 3.13 in a region $\mathcal{V}$ can be represented in $\mathcal{V}$ by an integral of propagating plane waves. Another representation is attributed to Weyl [1919], who showed that a spherical wave can be represented in any half-space not containing the point source by an integral of both propagating and evanescent plane waves. Expansion of the latter type (Weyl's expansion) are usually called *angular spectrum* expansions, as in Mandel and Wolf [1995, pp. 109-127].

Expansion of Whittaker's and Weyl's form are similar, in the sense that both are based on plane waves basis solutions, but they differ in some respect. Indeed, Whittaker's expansion contains only propagating plane waves, thus at any observation point it includes plane waves propagating in all directions, both incoming and outgoing with respect to the sound source. On the other hand, Weyl's expansion includes homogeneous plane waves propagating only in the outgoing half-space. It has been shown in Nieto-Vesperinas [1988] that the two representations are equivalent in their common domain of validity, in the sense that the Whittaker expansion of plane waves propagating towards the sound source is equal to the contribution of evanescent waves in Weyl's expansion.

4.1.1 Whittaker's Representation

Consider the acoustic field of a propagating plane wave of the form

$$P(\mathbf{r},\omega) = e^{j\langle \mathbf{k},\mathbf{r}\rangle}, \quad \mathbf{k}\in\mathbb{R}^3, \quad \|\mathbf{k}\| = \frac{\omega}{c}. \tag{4.2}$$

Upon varying $\mathbf{k}$ in $\mathbb{R}^3$, with the constraint $\|\mathbf{k}\| = \frac{\omega}{c}$ dictated by the dispersion relation, one obtains a complete set of basis functions over which an arbitrary acoustic field may be decomposed. The Whittaker expansion can thus be formulated as an inverse multi-dimensional Fourier transform with respect to the spatial variable $\mathbf{r}$ (cfr. Def. 3), i.e.

$$P(\mathbf{r},\omega) = \left(\frac{1}{2\pi}\right)^3 \iiint_{\mathcal{D}} \tilde{P}(\mathbf{k}) e^{j\langle \mathbf{k},\mathbf{r}\rangle}\, d^3\mathbf{r}, \quad \mathcal{D} = \left\{\mathbf{k}\in\mathbb{R}^3 : \|\mathbf{k}\| = \frac{\omega}{c}\right\}, \tag{4.3}$$

where $\mathbf{k}$ plays the role of a spatial frequency and the function $\tilde{P}(\mathbf{k})$ encodes the amplitude and phase of each plane wave contribution to the integral. We observe that the domain of integration in (4.3) is a sphere in $\mathbb{R}^3$ with radius ω/c; this peculiar domain of integration results from the dispersion relation, which dictates $\|\mathbf{k}\| = \omega/c$ in order to obtain a physically-valid acoustic field. The function $\tilde{P}(\mathbf{k})$ can be computed from the knowledge of the acoustic field $P(\mathbf{r},\omega)$ starting from the multi-dimensional Fourier transform of $P(\mathbf{r},\omega)$ with respect to $\mathbf{r}$

$$\tilde{P}(\boldsymbol{\xi}) = \iiint_{\mathbb{R}^3} P(\mathbf{r},\omega) e^{-j\langle \boldsymbol{\xi},\mathbf{r}\rangle}\, d^3\boldsymbol{\xi} \tag{4.4}$$

and then restricting $\tilde{P}(\boldsymbol{\xi})$ only to those values of $\boldsymbol{\xi}$ that satisfy the dispersion relation

$$\tilde{P}(\mathbf{k}) = \tilde{P}(\boldsymbol{\xi})|_{\|\boldsymbol{\xi}\|=\|\mathbf{k}\|=\frac{\omega}{c}}. \tag{4.5}$$

The geometric interpretation of (4.3) provides useful insights on the intuition behind Whittaker expansion. Upon factorizing the wavenumber vector $\mathbf{k}$ as

$$\mathbf{k} = \frac{\omega}{c}\hat{\mathbf{k}}, \tag{4.6}$$

where $\hat{\mathbf{k}}$ is the unit vector in the direction of $\mathbf{k}$, which can be written in terms of its Cartesian components as

$$\begin{aligned} \hat{k}_x &= \kappa \sin(\theta)\cos(\phi), \\ \hat{k}_y &= \kappa \sin(\theta)\sin(\phi), \\ \hat{k}_z &= \kappa \cos(\theta), \end{aligned} \tag{4.7}$$

θ and ϕ being the co-elevation and azimuth angles in the spherical reference frame, respectively, and $\kappa = 1$ to satisfy the dispersion relation (3.17). In order to substitute the angular factorization (4.7) into the Whittaker expansion, one has to perform a change of variables in the integration (4.3) from $\mathbf{k} = (\omega/c)[\hat{k}_x, \hat{k}_y, \hat{k}_z]^T$ to θ, ϕ, κ. The Jacobian matrix reads

$$\left|\frac{\partial(\hat{k}_x, \hat{k}_y, \hat{k}_z)}{\partial(\kappa, \theta, \phi)}\right| = \begin{vmatrix} \partial\hat{k}_x/\partial\kappa & \partial\hat{k}_x/\partial\theta & \partial\hat{k}_x/\partial\phi \\ \partial\hat{k}_y/\partial\kappa & \partial\hat{k}_y/\partial\theta & \partial\hat{k}_y/\partial\phi \\ \partial\hat{k}_z/\partial\kappa & \partial\hat{k}_z/\partial\theta & \partial\hat{k}_z/\partial\phi \end{vmatrix} = \kappa^2 \sin(\theta). \tag{4.8}$$

Setting $\kappa = 1$ to satisfy the dispersion relation and upon performing the change of variables using the Jacobian in (4.8), one obtains a different form of Whittaker's expansion

$$P(\mathbf{r}, \omega) = \left(\frac{\omega}{c}\right)^3 \iint_{\mathcal{S}} \tilde{P}(\theta, \phi, \omega) e^{j\frac{\omega}{c}(x\sin(\theta)\cos(\phi) + y\sin(\theta)\sin(\phi) + z\cos(\theta))} \sin(\theta)\, d\theta\, d\phi, \tag{4.9}$$

with the domain of integration being $\mathcal{S} = \{\theta \in [0, \pi], \phi \in [0, 2\pi)\}$. The interpretation of (4.9) is straightforward and intuitive. Equation (4.9) states that an arbitrary acoustic field can be represented as a superposition of propagating plane waves in directions $\theta \in [0, \pi]$, $\phi \in [0, 2\pi)$, while the function $\tilde{P}(\theta, \phi, \omega)$ encodes the magnitude and phase for each plane wave in direction θ, ϕ at temporal frequency ω. The function $\tilde{P}(\theta, \phi, \omega)$ is referred to in the literature as *Herglotz density* [Colton and Kress, 1992, p. 54]. It is remarkable that the representations in (4.3) and (4.9) are valid in a whole source free region $\mathcal{V}$ [Sherman, 1969] and $\tilde{P}$ is spatially invariant, in the sense that it does not depend on the observation position $\mathbf{r}$.

The notion of the Herglotz density is of great importance in this thesis. As it will be clear in next chapters, it proves it provide the theoretical foundation for several of the techniques presented in this thesis.

4.1.2 Weyl's Representation

Consider a source distribution

$$Q(\mathbf{r}, \omega) \neq 0, \quad \forall \mathbf{r} \in \mathcal{Q}. \tag{4.10}$$

The acoustic field generated by this source distribution satisfies the inhomogeneous Helmholtz equation (3.15) and can be constructed by means of Green's functions as

$$P(\mathbf{r}, \omega) = \int_{\mathcal{Q}} Q(\mathbf{r}', \omega) G(\mathbf{r}|\mathbf{r}', \omega)\, d^3\mathbf{r}'. \tag{4.11}$$

Weyl's identity [Williams, 1999, p. 35] states that the acoustic field of a point source can be expanded as an integral of both propagating and evanescent plane waves[1] [Mandel and Wolf, 1995, p. 123, Eq. 3.2-62]

$$G(\mathbf{r}|\mathbf{r}',\omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|} = -\frac{j(\omega/c)}{2\pi}\iint_{\mathbb{R}^2}\frac{e^{j(k_x(x-x')+k_y(y-y')+k_z|z-z'|)}}{k_z}\,dk_x\,dk_y, \quad (4.12)$$

where

$$k_z = \begin{cases}\sqrt{(\omega/c)^2-k_x^2-k_y^2}, & \text{for } k_x^2+k_y^2 \leq (\omega/c)^2\\ -j\sqrt{k_x^2+k_y^2-(\omega/c)^2}, & \text{for } k_x^2+k_y^2 > (\omega/c)^2\end{cases} \quad (4.13)$$

and the modulus $|z-z'|$ accounts for both half spaces $z-z'>0$ and $z-z'<0$. It should be remarked that (4.12) is valid unless $z=z'$. Upon substituting (4.12) into (4.11) one obtains

$$P(\mathbf{r},\omega) = -\frac{j(\omega/c)}{2\pi}\int_{\mathcal{Q}} Q(\mathbf{r}',\omega)\left[\iint_{\mathbb{R}^2}\frac{e^{j(k_x(x-x')+k_y(y-y')+k_z|z-z'|)}}{k_z}\,dk_x\,dk_y\right]d^3\mathbf{r}'. \quad (4.14)$$

Considering only the half space $z-z'>0$, after exchanging the order of integration and bringing terms independent on $\mathbf{r}'$ out of the inner integral, one obtains

$$P(\mathbf{r},\omega) = -\frac{j(\omega/c)}{2\pi}\iint_{\mathbb{R}^2} e^{j\langle\mathbf{k},\mathbf{r}\rangle}\left[\int_{\mathcal{Q}}\frac{Q(\mathbf{r}',\omega)}{k_z}e^{-j\langle\mathbf{k},\mathbf{r}'\rangle}\,d^3\mathbf{r}'\right]dk_x\,dk_y, \quad (4.15)$$

where the terms in square brackets in (4.15) is recognized as a multi-dimensional Fourier transform of $Q(\mathbf{r}',\omega)/k_z$, restricted to the domain $\mathcal{Q}$ where the sound sources are active. Equation (4.15) is the Weyl's expansion of an arbitrary acoustic field $P(\mathbf{r},\omega)$ and it involves both propagating and evanescent plane waves. The integral in square brackets is known as *angular spectrum* [Mandel and Wolf, 1995, p. 122], and it exhibits a discontinuity for $k_z=0$, i.e. for $(\omega/c)^2 = k_x^2+k_y^2$. The latter constraint represents a circle in the (k_x,k_y) plane where the nature of plane waves changes: propagating plane waves are encoded within the circle, while evanescent plane waves are outside the circle. It is shown in Mandel and Wolf [1995, pp. 122-123] that the singularity in (4.15) is integrable except at the position $\mathbf{r}'$ of the point source.

Also in the case of the Weyl's expansion, a more intuitive geometrical view can be derived, starting from the change of variables

$$\begin{aligned} k_x &= \frac{\omega}{c}\sin(\alpha)\cos(\beta),\\ k_y &= \frac{\omega}{c}\sin(\alpha)\sin(\beta),\\ k_z &= \frac{\omega}{c}\cos(\alpha). \end{aligned} \quad (4.16)$$

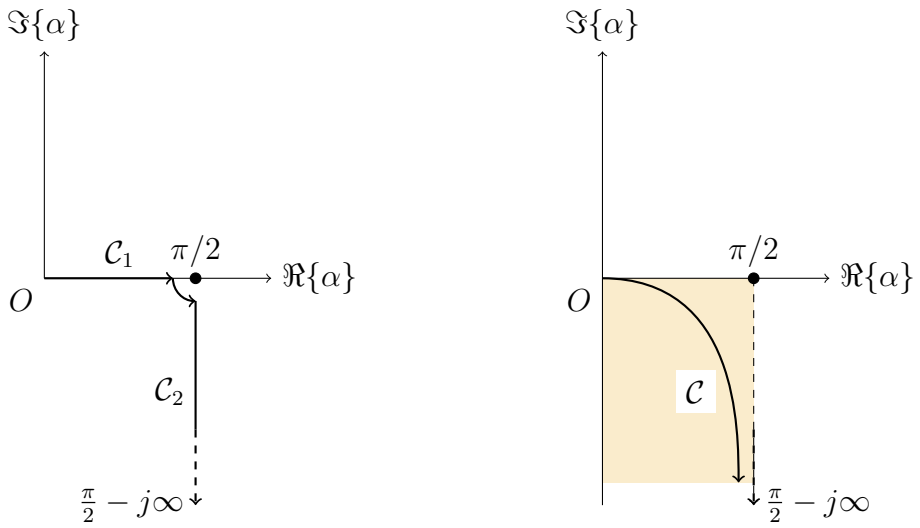
[1] In this thesis, the field of an outgoing spherical wave is described by

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

while in Williams [1999] a different convention is adopted, leading to the functional form

$$P(\mathbf{r},\omega) = \frac{e^{j\frac{\omega}{c}\|\mathbf{r}-\mathbf{r}'\|}}{4\pi\|\mathbf{r}-\mathbf{r}'\|}.$$

Due to this discrepancy, the form of Weyl's identity used in this thesis is taken from Mandel and Wolf [1995, p. 123] slightly different with respect to the one reported in Williams [1999, p. 35].



(a) Contours $\mathcal{C}_1$ and $\mathcal{C}_2$. **(b)** Contour $\mathcal{C}$.

Figure 4.1: The α-contours used in the Weyl representations of a spherical wave. In Fig. 4.1a each point on the horizontal segment $\mathcal{C}_1$ is associated with a homogeneous plane wave and each point on the vertical line $\mathcal{C}_2$ is associated with an evanescent plane wave. The curve $\mathcal{C}$ in Fig. 4.1b is the deformed contour. This figure has been redrawn from Mandel and Wolf [1995, p. 124, Fig. 3.14].

In the propagating regime, i.e. when $k_x^2 + k_y^2 \leq (\omega/c)^2$, the parameters α and β are the spherical angles related to the direction of propagation of the plane wave in the considered half space [Mandel and Wolf, 1995, p. 123, Eq. 3.2-64]

$$\alpha = \theta \in [0, \pi/2), \quad \text{and} \quad \beta = \phi \in [0, 2\pi). \tag{4.17}$$

In the evanescent case, i.e. when $k_x^2 + k_y^2 > (\omega/c)^2$, the wavenumber component k_z is imaginary, hence α is the complex angle [Mandel and Wolf, 1995, p. 123, Eq. 3.2-65]

$$\alpha = \frac{\pi}{2} + j\alpha', \quad -\infty < \alpha' < 0, \quad \text{and} \quad \beta = \phi \in [0, 2\pi). \tag{4.18}$$

Upon performing the change of variables (4.16) into (4.14) one obtains

$$P(\mathbf{r}, \omega) = -\frac{j(\omega/c)}{2\pi} \int_{\mathcal{Q}} Q(\mathbf{r}', \omega)$$
$$\left[\int_0^{2\pi} \int_{\mathcal{C}_1 + \mathcal{C}_2} e^{j\frac{\omega}{c}(\sin(\alpha)\cos(\beta)(x-x') + \sin(\alpha)\sin(\beta)(y-y') + \cos(\alpha)(z-z'))} \sin(\alpha)\, d\alpha\, d\beta \right] d^3\mathbf{r}', \tag{4.19}$$

where $\mathcal{C}_1$ denotes the portion of the α-contour associated to propagating plane waves, while $\mathcal{C}_2$ denotes the portion associated to evanescent waves. Figure 4.1a shows the two contours in the α plane. Thanks to the Cauchy theorem of integration in the complex plane [Rudin, 1987, p. 220], the contour $\mathcal{C}_1 + \mathcal{C}_2$ can be deformed into any contour that begins at the origin of the α plane and approaches asymptotically the point $\pi/2 - j\infty$. An example of such contour is the contour $\mathcal{C}$ depicted in Fig. 4.1b. Interchanging the order of integration in (4.19) and using the inner product notation, one obtains

$$P(\mathbf{r}, \omega) = -\frac{j(\omega/c)}{2\pi} \int_0^{2\pi} \int_{\mathcal{C}} e^{j\langle \mathbf{k}, \mathbf{r} \rangle} \tilde{Q}(\mathbf{k}) \sin(\alpha)\, d\alpha\, d\beta, \tag{4.20}$$

where the function $\tilde{Q}(\mathbf{k})$ has been introduced as the multi-dimensional Fourier transform of $Q(\mathbf{r}', \omega)$ with respect to spatial coordinates and restricted to the source domain $\mathcal{Q}$

$$\tilde{Q}(\mathbf{k}) = \int_{\mathcal{Q}} Q(\mathbf{r}', \omega) e^{-j\langle \mathbf{k}, \mathbf{r}' \rangle} \, d^3\mathbf{r}', \tag{4.21}$$

and the components of $\mathbf{k}$ are given in (4.16). Equation (4.20) represents an arbitrary sound field as the integral of plane waves propagating in one half space and of evanescent plane waves. The function $\tilde{Q}(\mathbf{k})$ encodes the amplitude and phase of each plane wave component.

Weyl's expansion is of great importance in this thesis. Particularly, results pertaining to Weyl's representation will be employed in next paragraphs to provide a solid physical foundation for the beam-based representation of acoustic field proposed in this thesis.

4.2 Spherical Wave Representation

The section reviews the spherical wave representation for acoustic fields, which arises from the choice of spherical waves (cfr. Sec. 3.3.2) as basis solutions to the Helmholtz equation. The spherical wave expansion describes an acoustic field as a sum of spherical harmonic waves weighed by suitable coefficients. The importance of this representation lies in the following considerations:

- the set of coefficients is infinite but discrete, not a continuum as in the plane wave representation;
- the series can often be truncated, in the sense that if one retains only the first terms, a good approximation to the acoustic field can still be obtained;
- the angular dependence of the acoustic field is completely described by spherical harmonics, thus it does not depend on temporal frequency.

Consider the acoustic field of a spherical wave as derived in Sec. 3.3.2

$$P(\mathbf{r}, \omega) = R(r) Y_l^m(\theta, \phi), \tag{4.22}$$

where the function $R(r)$ describes the functional dependency of the acoustic field on the observation distance $r = \|\mathbf{r}\|$

$$R(r) = R_1 j_l\left(\frac{\omega}{c} r\right) + R_2 y_l\left(\frac{\omega}{c} r\right), \text{ or} \tag{4.23}$$

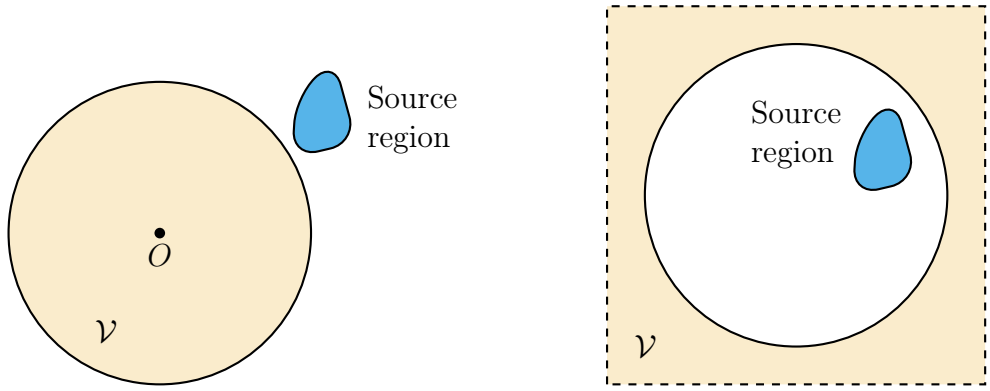
$$R(r) = R_3 h_l^{(1)}\left(\frac{\omega}{c} r\right) + R_4 h_l^{(2)}\left(\frac{\omega}{c} r\right). \tag{4.24}$$

We refer the reader to (3.44) and (3.45) for a discussion on the role played by all the terms in the two preceding equations. It suffice here to remark that the function $j_l(\cdot)$ is suitable to represent the acoustic field internal to a source distribution, while the external case is represented by the function $h_l^1(\cdot)$.

Any solution to the homogeneous Helmholtz equation can be written as the infinite sum of spherical waves [Williams, 1999, p. 186]

$$P(\mathbf{r}, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(A_{lm}(\omega) j_l\left(\frac{\omega}{c} r\right) + B_{lm}(\omega) \left(\frac{\omega}{c} r\right) \right) Y_l^m(\theta, \phi) \tag{4.25}$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(C_{lm}(\omega) h_l^{(1)}\left(\frac{\omega}{c} r\right) + D_{lm}(\omega) h_l^{(2)}\left(\frac{\omega}{c} r\right) \right) Y_l^m(\theta, \phi). \tag{4.26}$$



(a) Interior. **(b)** Exterior.

Figure 4.2: Interior and exterior problems. The gray area denotes the volume of interest $\mathcal{V}$, while the sound source region is colored in black.

In the following a discrimination between the two cases above mentioned cases is introduced, according to the spatial arrangement of acoustic sources with respect to the volume where the acoustic field representation is valid: the concepts of *interior* and *exterior* problems, according to the settings described in Williams [1999, p. 207 and p. 217], are introduced. On the one hand, interior problems consider acoustic fields in a volume that is free of sources, in the sense that all the acoustic sources and scatterers are located outside the volume of interest, as depicted in Fig. 4.2a. On the other hand, exterior problems consider the acoustic field in a volume that is exterior to the spatial region where sources and scatterers are located, as depicted in Fig. 4.2b. Exterior problems do not necessary extend to infinity [Ahrens, 2012, p. 30].

It is important to remark that there is a region of space where the acoustic field can not be represented as a summation of spherical waves. This is the ring-shaped region that share the same radial distance with respect to the source distribution. A similar limitation emerges in plane-wave representations. In particular, Weyl's representation is valid only in an half space free of sources, while Whittaker's representation is valid only in an arbitrarily-shaped source-free volume; both cases are analyzed in Lalor [1968], Devaney and Sherman [1973].

4.2.1 Exterior Problem

Consider the exterior problem as depicted in Fig. 4.2b. The radial dependency of the acoustic field, in this case, is conveniently described only by the function $h_l^{(1)}$, describing an outgoing wave [Williams, 1999, p. 207]. Thus, the expansion of a general external acoustic field is [Williams, 1999, p. 206]

$$P(\mathbf{r},\omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm}(\omega) h_l^{(1)}\left(\frac{\omega}{c} r\right) Y_l^m(\theta,\phi). \tag{4.27}$$

In this expansion, the acoustic field is completely determined by the coefficients $C_{lm}(\omega)$.

Assuming that the acoustic field is known on a sphere of radius a, the acoustic field in the entire external volume $\mathcal{V}$ is completely determined once the coefficients $C_{lm}(\omega)$ are known. After evaluating (4.27) on a sphere of radius a (i.e. $\|\mathbf{r}\| = 0$) and exploiting the

orthonormal property of spherical harmonics (cfr. the discussion on Eq. 3.40) one obtains [Williams, 1999, p. 207, Eq. 6.93]

$$C_{lm}(\omega) = \frac{1}{h_l^{(1)}\left(\frac{\omega}{c}r\right)} \int_0^{2\pi} \int_0^{\pi} P(a,\theta,\phi,\omega) Y_l^{-m}(\theta,\phi) \sin(\theta)\, d\theta\; d\phi, \tag{4.28}$$

where it has been explicitly indicated that the acoustic field is considered on a sphere of radius a, i.e.

$$P(a,\theta,\phi,\omega) = P(\mathbf{r},\omega)|_{\|\mathbf{r}\|=a}. \tag{4.29}$$

Upon substituting (4.28) into (4.27) one obtains the expression for the external acoustic field in all points satisfying $\|\mathbf{r}\| > a$ [Williams, 1999, p. 207, Eq. 6.94]

$$P(\mathbf{r},\omega) = \sum_{l=0}^{\infty} \frac{h_l^{(1)}\left(\frac{\omega}{c}r\right)}{h_l^{(1)}\left(\frac{\omega}{c}a\right)} \sum_{m=-l}^{l} Y_l^m(\theta,\phi) \int_0^{2\pi} \int_0^{\pi} P(a,\theta',\phi',\omega) Y_l^{-m}(\theta',\phi') \sin(\theta')\, d\theta'\; d\phi'. \tag{4.30}$$

4.2.2 Interior Problem

Consider now the interior problem as depicted in Fig. 4.2a. The radial dependence, in this case, must account for the finiteness of the acoustic field near the origin. Thus, spherical Bessel functions are chosen to describe the radial dependence [Williams, 1999, pp. 217-218]

$$P(\mathbf{r},\omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}(\omega) j_l\left(\frac{\omega}{c}r\right) Y_l^m(\theta,\phi). \tag{4.31}$$

The expansion coefficients $A_{lm}(\omega)$ can be determined from the knowledge of the acoustic field on a sphere of radius b, following a route analogous to the external case, yielding [Williams, 1999, p. 218, Eq. 6.140]

$$A_{lm}(\omega) = \frac{1}{j_l\left(\frac{\omega}{c}b\right)} \int_0^{2\pi} \int_0^{\pi} P(b,\theta,\phi) Y_l^{-m}(\theta,\phi) \sin(\theta)\, d\theta\; d\phi. \tag{4.32}$$

Inserting (4.32) into (4.31) one obtains an expression for the acoustic field at $\|\mathbf{r}\| < b$ given the knowledge of the field on a sphere of radius b [Williams, 1999, p. 218, Eq. 6-142]

$$P(\mathbf{r},\omega) = \sum_{l=0}^{\infty} \frac{j_l\left(\frac{\omega}{c}r\right)}{j_l\left(\frac{\omega}{c}b\right)} \sum_{m=-l}^{l} Y_l^m(\theta,\phi) \int_0^{2\pi} \int_0^{\pi} P(b,\theta',\phi',\omega) Y_l^{-m}(\theta',\phi')\, d\theta'\; d\phi'. \tag{4.33}$$

4.2.3 Bandlimited Spherical Wave Representations

Both expansion (4.27) and (4.31) are characterized by an infinite number of coefficients. It has been shown in Kennedy et al. [2007] that an arbitrary acoustic field can be expressed as a truncated spherical wave series of the type

$$P_L(\mathbf{r},\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} C_{lm}(\omega) h_l^{(1)}\left(\frac{\omega}{c}r\right) Y_l^m(\theta,\phi) \tag{4.34}$$

for external fields and of the type

$$P_L(\mathbf{r},\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} A_{lm}(\omega) j_l\left(\frac{\omega}{c}r\right) Y_l^m(\theta,\phi). \tag{4.35}$$

for internal fields, and the error introduced by the truncation is upper bounded.

With more details, Kennedy et al. [2007, p. 2546, Theorem 1] state that an arbitrary acoustic field $P(\mathbf{r}, \omega)$ having representation (4.27) or (4.31) can be truncated to $l \leq L$ terms as in (4.34) or (4.35), and the normalized truncation error (as defined in Kennedy et al. [2007, p. 2546, Eq. 28]) is upper bounded by $0.16127e^{-\Delta}$ provided that

$$L = \left\lceil \frac{e}{2} \frac{\omega}{c} \|\mathbf{r}\| \right\rceil + \Delta, \quad \Delta \in \mathbb{Z}^+. \tag{4.36}$$

This result provides the theoretical foundation for the use of *bandlimited* (i.e. truncated) spherical wave expansions, thus enabling their use in practical contexts. In this scenario, the acoustic field is compactly described by L^2 coefficients.

4.3 Cylindrical Wave Representation

The adoption of cylindrical waves as basis solutions to the Helmholtz equation leads to the cylindrical wave representation for acoustic field. The importance of this representation lies in the fact that cylindrical waves are a convenient wave object to describe height-invariant acoustic field, thus the expansion in terms of cylindrical waves is usually adopted when one intends to simplify three-dimensional problems considering only a two-dimensional plane, and assuming the acoustic field to be invariant with respect to the neglected spatial direction. Similarly to the spherical wave representation, coefficients in the cylindrical wave representation are infinite but discrete, and the series can be truncated to approximate the acoustic field with arbitrary accuracy.

Consider the acoustic field of a cylindrical wave, as derived in 3.3.3

$$P(\mathbf{r}, \omega) = R(\rho)Z(z)e^{jm\phi}, \tag{4.37}$$

where $\rho = \sqrt{x^2 + y^2}$ and the function $R(\rho)$ describes the radial dependency of the acoustic field

$$R(\rho) = R_1 J_m(k_\rho \rho) + R_2 Y_m(k_\rho \rho) \tag{4.38}$$

$$R(\rho) = R_3 H_m^{(1)}(k_\rho \rho) + R_4 H_m^{(2)}(k_\rho \rho), \tag{4.39}$$

being $k_\rho = \sqrt{(\omega/c)^2 - k_z^2}$. Arbitrary solutions to the homogeneous Helmholtz equations can be written as the infinite sum of cylindrical waves

$$P(\mathbf{r}, \omega) = \sum_{m=-\infty}^{\infty} e^{jm\phi} \frac{1}{2\pi} \int_{\mathbb{R}} \left[A_m(k_z, \omega) e^{jk_z z} J_m(k_\rho \rho) + B_m(k_z, \omega) e^{jk_z z} Y_m(k_\rho \rho) \right] dk_z \tag{4.40}$$

$$= \sum_{m=-\infty}^{\infty} e^{jm\phi} \frac{1}{2\pi} \int_{\mathbb{R}} \left[C_m(k_z, \omega) e^{jk_z z} H_m^{(1)}(k_\rho \rho) + D_m(k_z, \omega) e^{jk_z z} H_m^{(2)}(k_\rho \rho) \right] dk_z. \tag{4.41}$$

Both expressions in (4.40) and (4.41) can be recognized as a mono-dimensional inverse Fourier transform on k_z and a Fourier series in ϕ.

The cylindrical wave expansion specialized for external problems includes only Hankel functions of first kind and reads [Williams, 1999, p. 124, Eq. 4.50]

$$p(\mathbf{r}, \omega) = \sum_{n=-\infty}^{\infty} e^{jm\phi} \frac{1}{2\pi} \int_{\mathbb{R}} C_m(k_z, \omega) e^{jk_z z} H_m^{(1)}(k_\rho \rho) \, dk_z. \tag{4.42}$$

On the other hand, solutions for internal problems are better described by Bessel functions [Williams, 1999, p. 123, eq. 4.49]

$$P(\mathbf{r},\omega) = \sum_{m=-\infty}^{\infty} e^{jm\phi} \frac{1}{2\pi} \int_{\mathbb{R}} A_m(k_z,\omega) e^{jk_z z} J_m(k_\rho \rho)\, dk_z. \tag{4.43}$$

In this thesis, expansions (4.42) and (4.43) will not be used directly. Instead, cylindrical waves expansions will be used to build approximations to three-dimensional problems when all microphones or loudspeakers are aligned on a single plane. In this scenario it is customary to consider the acoustic field to be invariant on one spatial coordinate, say z, and then focus on a single plane, $z = 0$ for convenience. This restriction prevents the presence of acoustic waves coming from below or above the plane $z = 0$, thus k_z is constrained to zero.

With these restrictions, (4.42) and (4.44) become

$$p(\mathbf{r},\omega) = \sum_{n=-\infty}^{\infty} C_m(\omega) H_m^{(1)}\left(\frac{\omega}{c}\rho\right) e^{jm\phi}. \tag{4.44}$$

and

$$P(\mathbf{r},\omega) = \sum_{m=-\infty}^{\infty} A_m(\omega) J_m\left(\frac{\omega}{c}\rho\right) e^{jm\phi}, \tag{4.45}$$

respectively. Inversion formulas for the computation of the expansion coefficients $C_m(\omega)$ or $A_m(\omega)$ will be discussed later in this thesis, in relation to specific applicative settings. The representations (4.44) and (4.45) are usually referred to as *circular harmonic expansions*.

4.4 Beam-Based Representation of Acoustic Fields

This section introduces the beam-based representation of acoustic field that will be exploited for acoustic signal processing purposes in Part III of this thesis. The beam-based representation is here introduced in a slightly different way with respect to the other representations presented in this chapter. Indeed, for the sake of highlighting the peculiarities of the beam-based representation in comparison with the conventional representations previously examined, a derivation of the beam-based representation is proposed starting from the plane wave representation; from this starting point, the discussion on the frame-like nature of the beam-based representations follows smoothly.

The goal of this section is to obtain a frame-like representation (cfr. 2.4) for the acoustic field, expressed as a summation of beam wave objects of the type introduced in Sec. 3.8. For this purpose, we define the four-dimensional space of spatial positions $\mathbf{r}' = [x, y, 0]^T$ (i.e. the origins of the beam objects) and beam axes $\hat{\mathbf{k}}$ as

$$\mathbf{X} = (\mathbf{r}', \hat{\mathbf{k}}). \tag{4.46}$$

Consider a Cartesian grid on $\mathbf{X}$, where the sampling intervals on x and y axes are equal and given by $\bar{x}$. Consider also a uniform sampling in the $\hat{\mathbf{k}}$ plane, where the sampling interval is denoted by $\bar{k}$. With this notation, a point in the grid is denoted by

$$\mathbf{X}_{\boldsymbol{\mu}} = (\mathbf{r}'_{\mathbf{m}}, \hat{\mathbf{k}}_{\mathbf{n}}) = (m_1\bar{x}, m_2\bar{x}, n_1\bar{k}, n_2\bar{k}), \tag{4.47}$$

where the vector index notation $\boldsymbol{\mu} = (\mathbf{m}, \mathbf{n}) = (m_1, m_2, n_1, n_2)$ is used. As proposed in Shlivinski et al. [2004, p. 2047], a reference temporal frequency $\bar{\omega}$ is chosen to be greater

than the maximum temporal frequency of interest, i.e. $\bar{\omega} > \omega_{\max}$; then, the discrete space $\mathbf{X}_{\boldsymbol{\mu}}$ is chosen to be critically complete at $\omega = \bar{\omega}$, i.e.

$$\frac{\bar{\omega}}{c}\bar{k}\bar{x} = 2\pi. \tag{4.48}$$

The same grid is used at all temporal frequencies $\omega < \bar{\omega}$ of interest; for a specific temporal frequency ω, the overcompleteness parameter is given by

$$\nu(\omega) = \frac{\omega}{\bar{\omega}} < 1. \tag{4.49}$$

Thus, a frequency independent set of beam origins and axes is obtained by scaling the overcompleteness parameter ν with frequency, such that (4.49) is satisfied.

4.4.1 General Frame Expansion

Consider a two-dimensional window function and its dual, obtained by Cartesian multiplication of the corresponding one-dimensional functions, i.e. [Shlivinski et al., 2004, p. 2047, Eq. 22a-b]

$$\varphi_{\boldsymbol{\mu}}(\mathbf{r}', \omega) = \varphi(\mathbf{r}' - \mathbf{r}'_{\mathbf{m}}, \omega) e^{j\frac{\omega}{c}\langle \hat{\mathbf{k}}_{\mathbf{n}}, \mathbf{r}' - \mathbf{r}'_{\mathbf{m}} \rangle}, \tag{4.50}$$

$$\tilde{\varphi}_{\boldsymbol{\mu}}(\mathbf{r}', \omega) = \tilde{\varphi}(\mathbf{r}' - \mathbf{r}'_{\mathbf{m}}, \omega) e^{j\frac{\omega}{c}\langle \hat{\mathbf{k}}_{\mathbf{n}}, \mathbf{r}' - \mathbf{r}'_{\mathbf{m}} \rangle}. \tag{4.51}$$

An arbitrary acoustic field $P(\mathbf{r}', \omega)$ in the $z = 0$ plane can be represented by a Gabor frame as

$$P(\mathbf{r}', \omega) = \sum_{\boldsymbol{\mu}} a_{\boldsymbol{\mu}}(\omega) \varphi(\mathbf{r}', \omega), \tag{4.52}$$

where the frame expansion coefficients $a_{\boldsymbol{\mu}}(\omega)$ are given by the projection of the field onto the dual frame elements

$$a_{\boldsymbol{\mu}}(\omega) = \langle P(\mathbf{r}', \omega), \tilde{\varphi}_{\boldsymbol{\mu}}(\omega) \rangle = \iint P(\mathbf{r}', \omega) \tilde{\varphi}^*(\mathbf{r}' - \mathbf{r}'_{\mathbf{m}}, \omega) e^{-j\frac{\omega}{c}\langle \hat{\mathbf{k}}_{\mathbf{n}}, \mathbf{r}' - \mathbf{r}'_{\mathbf{m}} \rangle} \, d^2\mathbf{r}'. \tag{4.53}$$

The frame expansion in (4.53) is similar to the multidimensional extension of the local Fourier transform in Def. 8: this consideration enables the identification of the coefficients $a_{\boldsymbol{\mu}}(\omega)$ with the local plane wave spectrum of $P(\mathbf{r}', \omega)$ sampled at grid points $\mathbf{X}_{\boldsymbol{\mu}}$.

4.4.2 Gaussian Beam Expansion

The choice of Gaussian windows of the type (2.70) as prototype frame elements leads to several valuable properties, listed in Shlivinski et al. [2004] and here reported.

- Gaussian windows can be parametrized by the overcompleteness parameter (4.49) in order to provide the minimum frame ratio for all temporal frequencies of interest.
- Gaussian windows determine Gaussian beams of the form (3.133) as propagating elements.

Consider a two-dimensional Gaussian window of the form

$$\varphi(\mathbf{r}', \omega) = e^{-\frac{\omega}{c}\frac{\|\mathbf{r}'\|^2}{2b}}, \tag{4.54}$$

where the parameter b denotes the width of the Gaussian window. In the rest of this work, the window function (4.54) is employed as frame element for the representation of acoustic field.

The acoustic field radiated by a Gaussian window of the type (4.54) is obtained from (4.52) upon substituting $\varphi(\mathbf{r}', \omega)$ with a propagator $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$, which describes the acoustic field radiated by through a Gaussian window on the plane $z = 0$, yielding

$$P(\mathbf{r}, \omega) = \sum_{\boldsymbol{\mu}} a_{\boldsymbol{\mu}}(\omega) B_{\boldsymbol{\mu}}(\mathbf{r}, \omega). \tag{4.55}$$

The propagator $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$ can be conveniently represented by its Weyl's plane wave expansion (c.f.r. (4.11)), thus obtaining the form reported in Shlivinski et al. [2004, p. 2047, Eq. 26]

$$B_{\boldsymbol{\mu}}(\mathbf{r}, \omega) = \left(\frac{\omega/c}{2\pi}\right)^2 \iint_{\mathbb{R}^2} \Phi_{\boldsymbol{\mu}}(\hat{\mathbf{k}}, \omega) e^{j\frac{\omega}{c}\langle \hat{\mathbf{k}}, \mathbf{r}' \rangle + jk_z z}\, dk_x\, dk_y, \tag{4.56}$$

being

$$\Phi_{\boldsymbol{\mu}}(\hat{\mathbf{k}}, \omega) = \Phi(\hat{\mathbf{k}} - \mathbf{n}\bar{k}) \tag{4.57}$$

the Fourier transform of $\varphi_{\boldsymbol{\mu}}$ with respect to space, and $k_z = \sqrt{1 - \langle \hat{\mathbf{k}}, \hat{\mathbf{k}} \rangle}$.

Upon substituting the spatial Fourier transform of (4.54) into (4.56) and employing saddle point integration, as detailed in Melamed [1997], one discovers that the propagator $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$ has the form of the Gaussian beam discussed in Sec. 3.8. However, the Gaussian beam (3.133) has its propagation axis fixed to the z axis. In the case of $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$, on the other hand, the beam axis is made variable to cover arbitrary beam directions. The beam $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$ emerges from the point $\mathbf{r}'_{\mathbf{m}}$ in direction $(\theta_{\mathbf{n}}, \phi_{\mathbf{n}})$ determined by the wavenumber vector as

$$\hat{\mathbf{k}}_{\mathbf{n}} = \sin(\theta_n) \left[\cos(\phi_n), \sin(\phi_n)\right]^T. \tag{4.58}$$

The analytical expression for $B_{\boldsymbol{\mu}}(\mathbf{r}, \omega)$ is conveniently expressed in the beam coordinates $(x_{b_{\boldsymbol{\mu}}}, y_{b_{\boldsymbol{\mu}}}, z_{b_{\boldsymbol{\mu}}})$, where the coordinates $x_{b_{\boldsymbol{\mu}}}$ and $y_{b_{\boldsymbol{\mu}}}$ denote the transverse direction with respect to the beam axis, and $z_{b_{\boldsymbol{\mu}}}$ is the distance along the beam axis. Following Shlivinski et al. [2004, p. 2048, Eq. 28], the beam coordinates are related to the reference coordinates (x, y, z) by

$$\begin{bmatrix} x_{b_{\boldsymbol{\mu}}} \\ y_{b_{\boldsymbol{\mu}}} \\ z_{b_{\boldsymbol{\mu}}} \end{bmatrix} = \begin{bmatrix} \cos(\theta_{\mathbf{n}})\cos(\phi_{\mathbf{n}}) & \cos(\theta_{\mathbf{n}})\sin(\phi_{\mathbf{n}}) & -\sin(\theta_{\mathbf{n}}) \\ -\sin(\phi_{\mathbf{n}}) & \cos(\phi_{\mathbf{n}})0 & \\ \sin(\theta_{\mathbf{n}})\cos(\phi_{\mathbf{n}}) & \sin(\theta_{\mathbf{n}})\sin(\phi_{\mathbf{n}}) & \cos(\theta_{\mathbf{n}}) \end{bmatrix} \begin{bmatrix} x - x_{\mathbf{m}} \\ y - y_{\mathbf{m}} \\ z \end{bmatrix}. \tag{4.59}$$

Using these coordinates, the Gaussian beam solution to (4.56) can be expressed as [Shlivinski et al., 2004, p. 2048, Eq. 29]

$$B(\mathbf{r}, \omega) \approx \sqrt{\frac{-jz_{c,x_{\boldsymbol{\mu}}}}{z_{b_{\boldsymbol{\mu}}} - jz_{c,x_{\boldsymbol{\mu}}}} \frac{-jz_{c,y_{\boldsymbol{\mu}}}}{z_{b_{\boldsymbol{\mu}}} - jz_{c,y_{\boldsymbol{\mu}}}}}\, e^{j\frac{\omega}{c} z_{b_{\boldsymbol{\mu}}} + \frac{1}{2}\frac{x_{b_{\boldsymbol{\mu}}}^2}{z_{b_{\boldsymbol{\mu}}} - jz_{c,x_{\boldsymbol{\mu}}}} \frac{1}{2}\frac{y_{b_{\boldsymbol{\mu}}}^2}{z_{b_{\boldsymbol{\mu}}} - jz_{c,y_{\boldsymbol{\mu}}}}}. \tag{4.60}$$

Equation (4.60) represents a Gaussian beam with waist at $z_{b_{\boldsymbol{\mu}}} = 0$ and principal axes $x_{b_{\boldsymbol{\mu}}}$, $y_{b_{\boldsymbol{\mu}}}$. The Rayleigh ranges are, in general, different on the two principal axes, i.e. $z_{c,x_{\boldsymbol{\mu}}}$ and $z_{c,y_{\boldsymbol{\mu}}}$ are not constrained to be equal; this situation is usually referred to in the optics literature as *astigmatism* and it is caused by the beam tilt that reduces the radius of the beam in the $x_{b_{\boldsymbol{\mu}}}$ direction by factor $\cos(\theta_{\mathbf{n}})$ [Shlivinski et al., 2004, p. 2048, Eq. 30], determining the Rayleigh ranges

$$z_{c,x_{\boldsymbol{\mu}}} = b\cos^2(\theta_{\mathbf{n}}), \quad z_{c,y_{\boldsymbol{\mu}}} = b. \tag{4.61}$$

Table 4.1: Beam parameters expressed in the beam coordinates (4.59).

Beam waist radius	$\sigma_{0,x_\mu} = \sqrt{\frac{z_{c,x_\mu}}{(\omega/c)}}$	$\sigma_{0,y_\mu} = \sqrt{\frac{z_{c,y_\mu}}{(\omega/c)}}$
Beam radius	$\sigma_{x_\mu}(z_{b_\mu}) = \sigma_{0,x_\mu}\sqrt{1+\left(\frac{z_{b_\mu}}{z_{c,x_\mu}}\right)^2}$	$\sigma_{y_\mu}(z_{b_\mu}) = \sigma_{0,y_\mu}\sqrt{1+\left(\frac{z_{b_\mu}}{z_{c,y_\mu}}\right)^2}$
Radius of curvature	$R_{x_\mu}(z_{b_\mu}) = z_{b_\mu} + \frac{z_{c,x_\mu}^2}{z_{b_\mu}}$	$R_{y_\mu}(z_{b_\mu}) = z_{b_\mu} + \frac{z_{c,y_\mu}^2}{z_{b_\mu}}$

Table 4.1 summarizes the beam parameters introduced in Sec. 3.8, each expressed in terms of the beam coordinates in (4.59).

In Chap. 7 we employ the beam-based representation of acoustic field as introduced in this chapter in order to provide a physical interpretation for the plenacoustic transform introduced in Sec. 7.5.

4.5 Summary

- The acoustic field representations of most widespread use in the acoustic signal processing literature are introduced in Sec. 4.1-4.3 starting from the basic solutions to the Helmholtz equation reviewed in Chap. 3. A novel representation borrowed from the optics literature and based on the concept of Gaussian beams is introduced in Sec. 4.4.

- Plane wave representations enable the description of an acoustic field as an integral expansion of plane wave; a representation involving only propagating plane waves (Whittaker's representation) is discussed along with another representation involving both propagating and evanescent plane waves (Weyl's representation). A geometrical interpretation is provided in both cases. Analysis and synthesis operations are described in both representations.

- The spherical wave representation enables the description of an acoustic field as a series of spherical waves; it is shown that the series can be truncated by retaining only the first terms, and that the error introduced by truncation is upper bounded. Analysis and synthesis operations based on spherical waves are described for both exterior and interior problems.

- The cylindrical wave representation enables the description of an acoustic field as a series of cylindrical waves. The importance of this representation is identified in the ease of modeling simplified problems where propagation on a single plane is considered. In this setting, synthesis operations for external and internal problems are introduced.

- The frame-like Gaussian beam representation enables the description of an acoustic field in terms of a discrete summation of tilted Gaussian beams. Exploiting the properties of frame-like representations introduced in Sec. 2.4 and in Sec. 2.4.1, the parameters of the representation are determined in a peculiar fashion. This peculiarity yields to a representation where the beam axis and the quality of the representation are independent on the temporal frequency.

This page intentionally left blank.

Part II

Fourier Analysis and Synthesis of Acoustic Fields

This page intetntionally left blank.

Chapter 5

Analysis of Acoustic Fields

This chapter presents novel techniques for the analysis of acoustic fields. Analysis is conceived here as the extraction of salient information from data acquired by a spatial distribution of microphones, arranged in space in regular geometries. Acoustic field analysis is a crucial task in many applicative scenarios. The applicative class considered in this chapter focuses on the measurement of acoustic properties of an environment. Indeed, in studying the acoustic properties of closed environments, scholars have introduced a variety of objective parameters (the interested reader is referred to Kuttruff [2009] for a comprehensive discussion of those parameters), commonly measured by processing microphone array data. Of particular interest is the study of *early reflections*, i.e. sound events that are sensed by a listener after being reflected by surfaces in the listening environment, e.g. floor, walls and ceiling.

The analogy between DoAs and spatial frequencies in plane wave representations provide the foundation of all the techniques presented in this chapter. Indeed, state-of-the-art tools from the literature on spectral analysis are here adapted to the analysis of spatial features of acoustic fields. With more details, the methodology adopted in this chapter is outlined in the following.

1. An arbitrary acoustic field is represented as an integral of propagating plane waves (according to Whittaker's representation (4.1.1)).

2. The integral representation is discretized to provide an algebraic formulation whose central feature is a series of complex exponential contributions.

3. Extract model parameters (in general, amplitude, phase, DoA and Time of Arrival (ToA) of each plane wave component) from acoustic field measurements performed by a microphone array.

A technique implementing the methodology outlined above is presented in this chapter. The technique is based on short-time Fourier analysis to retrieve time-varying information on the spatial frequency components of the acoustic field; based on this information, a statistical model is matched to the actual array data, as in Yardibi et al. [2010]. The proposed solution exhibits good robustness to modeling mismatches and to real-world recording conditions; its effectiveness is proved experimentally in applications of acoustic measurements both in acoustically conditioned environments and in unconditioned environments.

Spherical arrays represent the choice of election and have been widely studied in the literature. Spherical arrays of different flavors have been investigated by scholars, ranging from open arrays [Gover et al., 2002], rigid arrays (where microphones are placed

on the surface of a rigid sphere) [Abhayapala and Ward, 2002, Meyer and Elko, 2002] and hybrid configurations (e.g. a smaller rigid array and a concentric open array with larger radius [Parthy et al., 2009]). Here the focus is on the use of such techniques for acoustic measurements, where the goal is to estimate strength, DoA and ToA of plane wave components impinging on the spherical array. This is usually achieved considering short-time segments of the signals acquired by the array, estimating modal coefficients as outlined in Sec. 4.2.2 and then transforming the spherical wave representation to the plane wave representation. This task can be accomplished using several techniques; among the others we list MUSIC with time/frequency smoothing [Khaykin and Rafaely, 2012, Huleihel and Rafaely, 2013], sub-space pre-processing [Epain and Jin, 2013] and maximum likelihood estimation [Tervo et al., 2013].

Plane waves impinging on the microphone array carry a great deal of information on the acoustics of the environment. Indeed, each plane wave component can be associated to a specific acoustic event, namely the direct sound reaching the array along the direct path and early reflections. Early reflections can be regarded to as a set of spatio-temporal events occurring at distinct time instants and originating from distinct spatial locations. Since early reflections coming from walls and obstacles could represent a problem for the listener (in the sense that they distort the desired perception of a sound scene), researchers in acoustics and acoustic signal processing community have worked in the past on the accurate determination of the direction of arrival of early reflections. Many techniques have been proposed in the literature (e.g. in Gover et al. [2002, 2004], Khaykin and Rafaely [2012], Martellotta [2013] and references therein) based on *acoustic impulse responses*, i.e. measurements obtained by exciting the environment with a proper sound source, recording the room response with an omnidirectional microphone and then processing the recorded signal to obtain the impulse response [Müller and Massarani, 2001]. The determination of direction and strength of early reflections is a fundamental task in the design and optimization of architectural spaces from the acoustics standpoint: since basic geometric considerations enable to associate a reflection with the architectural element that originated it, the analysis of reflections allows to modify the environment in a thoughtful way. Furthermore, this task is important also in applications of room equalization and correction [Carini et al., 2012, Canclini et al., 2012, 2014b,a, Poletti et al., 2015], in the identification and control of industrial and aerospace noise sources [Gover, 2005] and for speech enhancement [Ribeiro et al., 2010].

The theoretical tools developed in the literature and reviewed in Sec. 4.2.2 could be used to estimate the parameters of plane waves and identify early reflections starting from these parameters. Indeed, in O'Donovan et al. [2008] a panoramic projection of the visual image of the enclosure is layered with the magnitude of directional impulse response. These responses are obtained through a beamforming based on the expansion of the acoustic field captured by the array in terms of spherical harmonics. In Farina et al. [2011, 2010] the previous approach is refined by using measured directional array transfer functions to deconvolve the acquired signals. The main issue of this refinement lies in the fact that a preliminary measurement session in an anechoic chamber is required, in order to determine the response of each microphone in the array to sound waves impinging from different directions.

The rest of the chapter is structured as follows. Sec. 5.1 presents the proposed analysis methodology, starting from the algebraic modeling of the underlying acoustic system (plane wave propagation and sphere scattering); then Sec. 5.2 presents the application of a state-of-the-art parameter estimation technique proposed in Stoica et al. [2011a] (borrowed from the literature on spectral analysis) to estimate the plane wave DoAs and strengths. After that, Sec. 5.3 shows experimental results that prove the effectiveness of the

proposed technique: first, results related to the validation in an acoustically conditioned environments are shown; then, results obtained in a challenging real world environment (the *Auditorium "Giovanni Arvedi"*, located in the *Museo del Violino* (Violin Museum), Cremona, Italy are shown. Results show that the proposed technique is able to estimate amplitude, strength and ToA of acoustic reflections with remarkable accuracy.

This chapter is focused on the analysis of three-dimensional acoustic field by means of a spherical microphone array. In particular, a methodology for estimating the temporal and spatial distribution of reflections in an enclosed space is described. Particular attention is devoted to the data visualization paradigm, chosen carefully to provide an accessible and intuitive source of information. The analysis is based on the impulse responses acquired by exciting the environment with an impulsive sound source, as in Müller and Massarani [2001].

The work presented here exploits the fact that the underlying model for array data, abstracting from the actual plane-wave model, assumes the general from of a spectral model, where the modeled function is conceived as a sum of complex exponential contributions. The use of such a spectral model paves the way to the generalization and adaptation of state-of-the-art spectral analysis techniques to estimate the model parameters. In particular, since in most real-world measurement cases one can not rely on much a-priori information, a non-parametric approach is presented here; the absence of parameters makes the proposed method robust to modeling mismatches and frees the user from the need of carefully tuning the parameters required by other techniques. This constitutes a great advantage with respect to other state of the art acoustic field analysis methodologies.

In this chapter a rigid spherical microphone array configuration is adopted. This is a widely used model in array processing with spherical microphone arrays (adopted e.g. in Meyer and Elko [2002], Rafaely [2004], Park and Rafaely [2005]). The directional imaging problem is formulated here as a spectral analysis problem, representing the sound field as a spectrum of propagating plane waves. A state-of-the art spectral analysis tool presented in Stoica et al. [2011a,b], based on matching the covariance matrix of array data to a prescribed model, is used to obtain an high resolution estimate of the plane waves expansion coefficients. Since the magnitude of these coefficients is related to the sound intensity coming from a specific direction [Fahy, 1995, Sec. 4.4], an estimate is obtained of the intensity of the reflections coming from a specific direction. Finally, the reflection intensity diagram is mapped on a panoramic visual image of the environment. The methodology presented here has already been published by the authors in Bianchi et al. [2015a].

5.1 Data Model

This section introduces the data model adopted throughout the rest of the paragraph. The considered scenario is that of an acoustic field generated by a single sound source in a reflective environment and acquired by a spherical microphone array. Under the assumption of far-field propagation (i.e. the excitation source and the reflectors are considered to be in the far field with respect to the microphone array, as in O'Donovan et al. [2008]), the reflections can be parametrized according to their angle of incidence on the microphone array.

The physical foundation of the proposed methodology is the Whittaker-type plane wave expansion (c.f.r. Sec. 4.1.1), in which the acoustic field is represented as the integral

of propagating plane waves

$$P(\mathbf{r},\omega) = \int_0^{2\pi}\int_0^{\pi} \tilde{P}(\theta,\phi,\omega)e^{j\langle \mathbf{k},\mathbf{r}\rangle}\sin(\theta)\,d\theta\;d\phi, \tag{5.1}$$

where the wavenumber vector is parametrized as $\mathbf{k} = [\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)]^T$, being $\theta \in [0,\pi]$ and $\phi \in [0,2\pi)$ the co-elevation angle and azimuth angle, respectively.

In the considered scenario, however, prevalent contributions to the acoustic field come from a specific set of discrete directions; hence, the integral equation (5.1) is simplified as

$$P(\mathbf{r},\omega) = \sum_{q=1}^{Q} e^{j\langle \mathbf{k}((\theta,\phi)_q,\omega),\mathbf{r}\rangle}\tilde{P}((\theta,\phi)_q,\omega), \tag{5.2}$$

where Q is the total number of reflections contributing to the field and the notation $(\theta,\phi)_q$ denotes the angles associated to the qth reflection; we denote by Γ the set of all reflections, i.e. $(\theta,\phi)_q \in \Gamma$. In principle, the model (5.2) could be directly used to estimate the magnitude of each reflection, i.e. $|\tilde{P}((\theta,\phi)_q,\omega)|$. However, this naive solution suffers from several problems that prevent its practical applicability. The first problem is the non-feasibility of employing a device with a continuous pressure sensitive spherical surface; thus, a discrete distribution of microphone capsules has to be adopted on the spherical surface itself. This operation introduces aliasing errors if the complexity of the observed acoustic field is greater than the one allowed by the adopted sampling schema. A discussion of aliasing phenomena in spherical microphone arrays is out of the scope of this thesis; we refer the interested reader to Rafaely [2005], Li and Duraiswami [2007]. Another source of errors that prevents the direct exploitation of the model in (5.2) is identified in the microphones self-noise [Rafaely, 2005]. The methodology presented in the following is intended to circumvent these limitations; indeed the proposed method is based on a parameter estimation technique that is known to enhance the resolution of conventional estimators [Stoica et al., 2011a], thus reducing the impact of aliased component on the estimate; moreover, noise is explicitly taken into account in the data model and its statistics is exploited in order to refine the estimate.

The data model used in this chapter is directly derived from (5.2) but modified to accommodate a short-time analysis setting. Indeed, the local (temporal) Fourier transform of the acoustic field can be written as

$$P(\mathbf{r},\omega,t) = \sum_{q=1}^{Q} e^{j\langle \mathbf{k}((\theta,\phi)_q,\omega,t),\mathbf{r}\rangle}\tilde{P}((\theta,\phi)_q,\omega,t), \tag{5.3}$$

where $\mathbf{k}((\theta,\phi)_q,\omega,t)$ is proportional to the plane wave DoA at time t and $\tilde{P}((\theta,\phi)_q,\omega,t)$ denotes the complex strength of the plane wave a time t. In particular, let $\mathbf{y}(t) \in \mathbb{R}^{M\times 1}$ be the impulse response vector acquired by a spherical microphone array composed by M microphones. The presented methodology is based on short-time segments of the acquired impulse responses, each weighed by a suitable window function. In this sense, let $\mathbf{y}(\omega_l,t_n) \in \mathbb{C}^{M\times 1}$ be the lth bin of the short-time discrete Fourier transform [Allen, 1977] (cfr. (2.78)) for the segment centered in t_n, i.e. $\omega_l = 2\pi f_l/F_s$, being f_l and F_s the temporal frequency and the temporal sampling frequency, respectively. The signal acquired by the microphone array is modeled as the sum of the acoustic field $P(\mathbf{r},\omega,t)$ and a noise component (that, in general, accommodates for the presence of microphone noise, sensor placement errors, etc.). In the following, the notation in (5.2) is adopted,

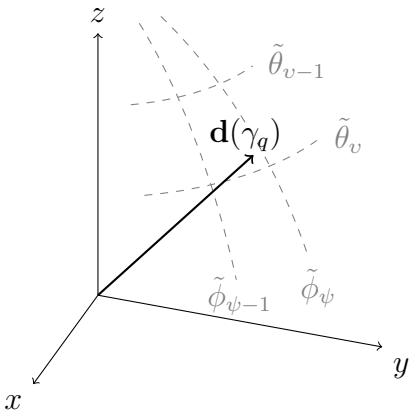


Figure 5.1: Angular grid adopted for the estimation of acoustic reflections.

denoting by $\mathbf{a}(\cdot) : \Gamma \to \mathbb{C}^{M\times 1}$ the propagation function that includes both plane wave propagation from direction γ_q and the scattering due to the rigid sphere.

$$\mathbf{y}(\omega_l, t_n) = \sum_{q=1}^{Q} \mathbf{a}(\gamma_q, \omega_l) s_q(\omega_l, t_n) + \mathbf{e}(\omega_l, t_n), \tag{5.4}$$

where $\mathbf{e}(\omega_l) \in \mathbb{C}^{M\times 1}$ is the additive noise term; $\gamma_q = (\theta, \phi)_q \in \Gamma$ and $s_q = \tilde{P}(\gamma_q, \omega) \in \mathbb{C}$ are the unknown parameters of the qth reflection, i.e. the DoA and the associated signal, respectively.

The function $\mathbf{a}(\gamma_q, \omega_l)$ encodes the modifications undertaken by a planar wave field component impinging on the microphone array from direction γ_q, taking into account the scattering operated by the rigid spherical surface. More formally, the mth component of $\mathbf{a}(\gamma_q, \omega_l)$ represents the pressure generated at the mth microphone due to an impulsive excitation coming from direction γ_q. Denoting by $\mathbf{r}_m \in \mathbb{R}^3$ the position vector relative to the mth microphone located on the surface of a rigid sphere of radius r, and by $\mathbf{d}_q \in \mathbb{R}^3$ the unit vector in the direction γ_q, authors in Zotkin et al. [2010, Eq. (10)] derived that

$$\{\mathbf{a}(\gamma_q, \omega_l)\}_m = \frac{i}{(r\omega_l/c)^2} \sum_{\mu=0}^{\infty} \frac{j^n(2\mu+1)\mathsf{P}_\mu(\langle \mathbf{r}_m, \mathbf{d}_q\rangle)}{h'_\mu(r\omega_l/c)}, \tag{5.5}$$

where $\mathsf{P}_\mu(\cdot)$ is the Legendre polynomial of degree μ [Olver, 2010, Chap. 18] ; $\langle \mathbf{r}_m, \mathbf{d}_q\rangle$ is the standard inner product in $\mathbb{R}^3$; and $h'_\mu(\cdot)$ is the first derivative of the spherical Hankel function of first kind and order μ [Olver, 2010, Sec. 10.47].

The goal of the presented methodology is to estimate the parameters γ_q and $|s_q|^2$ associated with each reflection observed in the nth frame of the acquired signals. However, the total number of reflections Q and their associated directions of arrival are not known a-priori. We adopt here a non-parametric estimation method on a predefined angular grid. Let $\{\tilde{\gamma}_\mathbf{u}\}$, $\mathbf{u} = (\upsilon, \psi) \in \mathbb{Z}^2$ denote the element of a grid that covers Γ and we assume that each γ_q is sufficiently close to a grid point, as in Fig. 5.1, i.e. $\gamma_q \approx \tilde{\gamma}_\mathbf{u}$; let also $U = \Upsilon\Psi$ be the total number of grid points, $\mathbf{a}_\mathbf{u}(\omega_l) = \mathbf{a}(\tilde{\gamma}_\mathbf{u}, \omega_l)$ and

$$\tilde{s}_\mathbf{u}(\omega_l, t_n) = \begin{cases} s_q(\omega_l, t_n), & \text{if } \tilde{\gamma}_\mathbf{u} = \gamma_q \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

Using this notation, the model in (5.4) can be rewritten as

$$\mathbf{y}(\omega_l, t_n) = \mathbf{A}(\omega_l)\mathbf{S}(\omega_l, t_n) + \mathbf{e}(\omega_l, t_m), \tag{5.7}$$

where $\mathbf{A}(\omega_l) = [\mathbf{a}_{(1,1)}(\omega_l), \ldots, \mathbf{a}_{(\Upsilon,\Psi)}(\omega_l)]$ and $\{\mathbf{S}(\omega_l, t_n)\}_{\mathbf{u}} = \tilde{s}_{\mathbf{u}}(\omega_l, t_n)$, so that $\mathbf{A}(\omega_l) \in \mathbb{C}^{M\times U}$ and $\mathbf{S}(\omega_l, t_n) \in \mathbb{C}^{U\times 1}$.

The following paragraphs show how to accurately estimate $|\tilde{s}_u(\omega_l, t_n)|^2$ given the array data, and how these estimates are associated to a panoramic visual image of the environment under test in order to provide an high-resolution reflection intensity map. One may observe that the estimation problem based on the data model in Eq. (5.7) is of widespread use in many application fields; in particular, it is widely studied in the context of spectral analysis. This fact enables to adapt to our problem solutions studied in other fields.

5.2 Parameter Estimation

This section briefly reviews the estimation method introduced in Stoica et al. [2011a,b] and it shows how this estimation method can be applied to the problem at hand. In order to simplify the notation, in this section the dependency of the data on ω_l and t_n is omitted.

Under the assumptions of uncorrelated noise

$$E[\mathbf{e}(\omega_l, t_n)\mathbf{e}^H(\omega_l, t_n)] = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_M^2 \end{bmatrix}, \tag{5.8}$$

and noise uncorrelated to sound sources, the covariance matrix of the array data can be written as [Stoica and Moses, 2004, Eq. (6.4.3)]

$$\mathbf{R} = E[\mathbf{y}\mathbf{y}^H] = \mathbf{A}^H\mathbf{\Sigma}\mathbf{A} + \mathbf{E}, \tag{5.9}$$

where the matrix $\mathbf{\Sigma}$ contains the power of individual signal components arranged on its diagonal, i.e. $\mathbf{\Sigma} = \mathrm{diag}([\sigma_1, \ldots, \sigma_U])$, being $\sigma_u = |\tilde{s}_u|^2$, and $\mathbf{E} = \mathrm{diag}([\epsilon_1, \ldots, \epsilon_M])$, ϵ_m denoting the noise variance at mth microphone. In the considered problem it is highly probable that a single impulse associated to an acoustic reflection is present in a frame, hence it can be safely assumed that the source signals are uncorrelated inside a temporal frame. However, in case multiple impulses were present inside the same temporal frame, in Stoica et al. [2011a] it is proved that the method reviewed here is robust to this assumption, thus the estimate is not impaired.

By defining

$$\check{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{I}_{M\times M} \end{bmatrix} \quad \text{and} \quad \check{\mathbf{\Sigma}} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0}_{U\times M} \\ \mathbf{0}_{M\times U} & \mathbf{E} \end{bmatrix}, \tag{5.10}$$

being $\mathbf{I}$ and $\mathbf{0}$ the identity matrix and the zero matrix, respectively. Equation (5.9) can be rewritten in the compact form

$$\mathbf{R} = \check{\mathbf{A}}^H\check{\mathbf{\Sigma}}\check{\mathbf{A}}, \quad \mathbf{R} \in \mathbb{C}^{W\times W}, \quad W = U + M. \tag{5.11}$$

In the following, $\check{\sigma}_w$ denotes the wth element on the diagonal of $\check{\mathbf{\Sigma}}$ and $\check{\mathbf{a}}_w$ denotes the wth column of $\check{\mathbf{A}}$.

The problem of estimating $\{\check{\sigma}_w\}_{w=1}^W$ is solved by fitting the modeled covariance matrix $\mathbf{R}$ to its sample estimate from array data $\mathbf{y}\mathbf{y}^H$, i.e. by finding the set $\{\check{\sigma}_w\}_{w=1}^W$ that satisfies [Stoica et al., 2011a, Eq. (19)]

$$\underset{\{\check{\sigma}_w\}}{\arg\min} \ \|\mathbf{R}^{-1/2}(\mathbf{y}\mathbf{y}^H - \mathbf{R})\|_F^2, \tag{5.12}$$

Algorithm 5.1 Parameter estimation algorithm.

$\check{\sigma}_w^{(0)} = |\check{\mathbf{a}}_w^H \mathbf{y}|^2 / \|\check{\mathbf{a}}_w\|^4$
$i = 1$
while $\|\boldsymbol{\Sigma}^{(i)} - \boldsymbol{\Sigma}^{(i-1)}\| / \|\boldsymbol{\Sigma}^{(i-1)}\| < \tau$ **do**
  $\mathbf{R} = \sum_{w=1}^{W} \check{\sigma}_w \check{\mathbf{a}}_w \check{\mathbf{a}}_w^H$
  $\mathbf{z} = \mathbf{R}^{-1}\mathbf{y}$
  $\zeta_w^{1/2} = \|\check{\mathbf{a}}_w\| / \|\mathbf{y}\|$
  $\xi_w = |\check{\mathbf{a}}_w^H \mathbf{z}|$
  $\rho = \sum_{w=1}^{W} \zeta_w^{1/2} \check{\sigma}_w^{(i-1)} \xi_w$
  $\check{\sigma}_w^{(i)} = \check{\sigma}_w^{(i-1)} \xi_w \rho / \zeta_w^{1/2}$
  $i = i + 1$
end while

where here $\|\cdot\|_F$ denotes the Frobenius norm for matrices. By expanding the cost function, the problem in (5.12) can be equivalently formulated as [Stoica et al., 2011a, Eq. (22)]

$$\underset{\{\check{\sigma}_w\}}{\arg\min} \ \operatorname{tr}(\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y}) + \sum_{w=1}^{U+M} h_w^2 \check{\sigma}_w, \tag{5.13}$$

where $h_w = \|\check{\mathbf{a}}_w\| / \|\mathbf{y}\|$ and $\operatorname{tr}(\cdot)$ denotes the matrix trace. The minimization problem in (5.13) is convex and has a global minimum, as shown in Stoica et al. [2011a, Sec. III-A].

An efficient iterative algorithm is derived in Stoica et al. [2011a, Sec. III-B] to solve (5.13) and it is summarized in Algorithm 5.1. Authors start by considering the modified problem

$$\underset{\{\check{\sigma}_w\},\mathbf{B}}{\arg\min} \ \operatorname{tr}(\mathbf{B}^H \check{\boldsymbol{\Sigma}} \mathbf{B}) + \sum_{w=1}^{U+M} h_w^2 \check{\sigma}_w \quad \text{s.t} \quad \check{\mathbf{A}}\mathbf{B} = \mathbf{y}\mathbf{y}^H. \tag{5.14}$$

The minimization over $\mathbf{B}$ for a fixed set $\{\check{\sigma}_w\}$ is given by $\hat{\mathbf{B}} = \check{\boldsymbol{\Sigma}} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{y}\mathbf{y}^H$ [Stoica et al., 2011a, Appendix A]. Substituting $\hat{\mathbf{B}}$ into (5.14) yields the original problem in (5.13). These considerations enable to conclude that the sets $\{\check{\sigma}_w\}$ obtained from (5.13) and (5.14) must be identical.

Upon defining $\mathbf{B} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_W]^T$, the problem (5.14) for a fixed set $\{\boldsymbol{\beta}_w\}$ can rewritten as

$$\underset{\{\check{\sigma}_w\}}{\arg\min} \ \sum_{w=1}^{W} \frac{\|\boldsymbol{\beta}_w\|^2}{\check{\sigma}_w} + \sum_{w=1}^{W} h_w^2 \check{\sigma}_w. \tag{5.15}$$

It is shown in Stoica et al. [2011a] that the minimizer of (5.15) is

$$\check{\sigma}_w = \frac{\|\boldsymbol{\beta}_w\|}{h_w}, \quad w = 1, \ldots, W. \tag{5.16}$$

Since the cost function in (5.14) is convex in both $\mathbf{B}$ and $\{\check{\sigma}_w\}$, a cyclic iterative minimization over $\mathbf{B}$ and $\{\check{\sigma}_w\}$ leads to a global minimum. The ith iteration involves the operations

$$\begin{aligned} \mathbf{B}^{(i)} &= \check{\boldsymbol{\Sigma}}^{(i-1)} \check{\mathbf{A}} \mathbf{R}^{-1} (i-1) \mathbf{y}\mathbf{y}^H \\ \check{\sigma}_w^{(i)} &= \|\boldsymbol{\beta}_w^{(i)}\| / h_w, \quad w = 1, \ldots, W \\ \mathbf{R}(i) &= \check{\mathbf{A}} \check{\boldsymbol{\Sigma}}^{(i)} \check{\mathbf{A}}^H. \end{aligned} \tag{5.17}$$

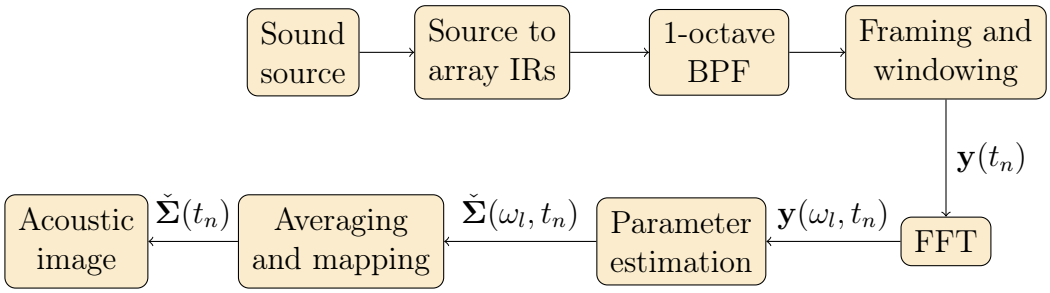


Figure 5.2: Block-diagram of the proposed imaging system for acoustic reflections.

Following Stoica et al. [2011a] for the initialization of (5.17), the beamforming estimator

$$\check{\sigma}_w^{(0)} = \frac{|\check{\mathbf{a}}_w^H \mathbf{y}|^2}{\|\check{\mathbf{a}}_w\|^4} \tag{5.18}$$

is used. The algorithm is stopped when the condition

$$\|\mathbf{\Sigma}^{(i)} - \mathbf{\Sigma}^{(i-1)}\| / \|\mathbf{\Sigma}^{(i-1)}\| < \tau \tag{5.19}$$

is satisfied for a given value of the threshold τ.

5.3 Experimental Validation

This section shows some experimental results of the proposed high-resolution imaging approach. For reference purposes, Fig. 5.2 shows the overall block diagram of the imaging system that has been implemented and that is subject to validation in this section.

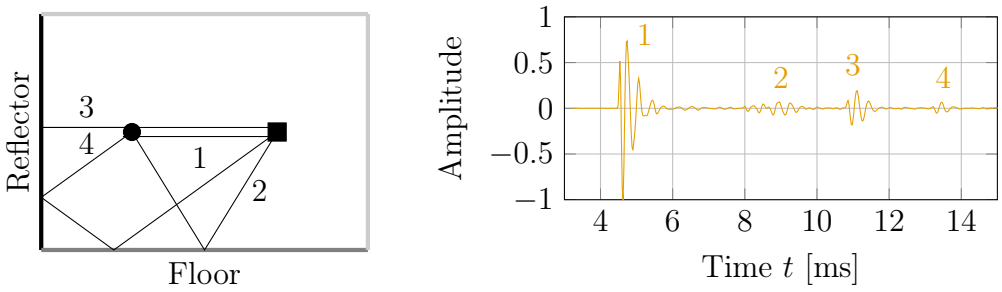
5.3.1 Setup

All the experiments are performed with *mh acoustics' Eigenmike*® spherical microphone array, composed of $M = 32$ capsules mounted on a rigid sphere, whose exact locations can be found in mh [2013]. The technique presented in 5.2 can be straightforwardly applied to any rigid spherical microphone array.

The angular grid $\{\tilde{\gamma}_{\mathbf{u}}\}_{\mathbf{u}=(1,1)}^{(\Upsilon,\Psi)}$ is designed to uniformly cover the spherical angular region of interest Γ. For this purpose, the angular axes θ and ϕ are uniformly sampled over Υ and Ψ points, respectively. In this setting, the double index $\mathbf{u}$ can be conveniently sorted as $u = (\upsilon - 1)\Psi + \psi$, with $\upsilon = 1, \ldots, \Upsilon$ and $\psi = 1, \ldots, \Psi$, and $u = 1, \ldots, U$.

In all the experiments, the sampling frequency is set to $F_s = 44.1\,\text{kHz}$ and frames of length 1.5 ms are considered. The impulse responses are measured using the exponential sine sweep technique [Farina, 2000]. The acquired impulse responses are processed with an octave pass-band filter centered at $f_c = 4\,\text{kHz}$, as recommended in Farina et al. [2011], for the purpose of a better identification of reflections. All the frequency bins in the passband of the filter concur to the generation of a single acoustic image, with the product of their geometric and harmonic means as suggested in Azimi-Sadjadi et al. [2004]

$$\check{\mathbf{\Sigma}}(t_n) = \frac{l_2 - l_1}{\sum_{l=l_1}^{l_2} \check{\mathbf{\Sigma}}^{-1}(\omega_l, t_n)} \cdot \left(\prod_{l=l_1}^{l_2} \check{\mathbf{\Sigma}}(\omega_l, t_n) \right)^{1/(l_2 - l_1)}, \tag{5.20}$$



(a) Geometry and main acoustic paths.

(b) Impulse response.

Figure 5.3: Placement of the spherical array in the controlled environment (Fig. 5.3a). Impulse response recorded by the first capsule (Fig. 5.3b); peaks are labeled according to the acoustic paths shown in Fig. 5.3a.

where the integers l_1 and l_2 are the bin indexes corresponding to the cutoff frequencies of the octave band-pass filter $\omega_{1,2} = 2\pi l_{1,2} F_s / L$. The threshold τ determining the convergence of Alg. 5.1 is set to 10^{-12}.

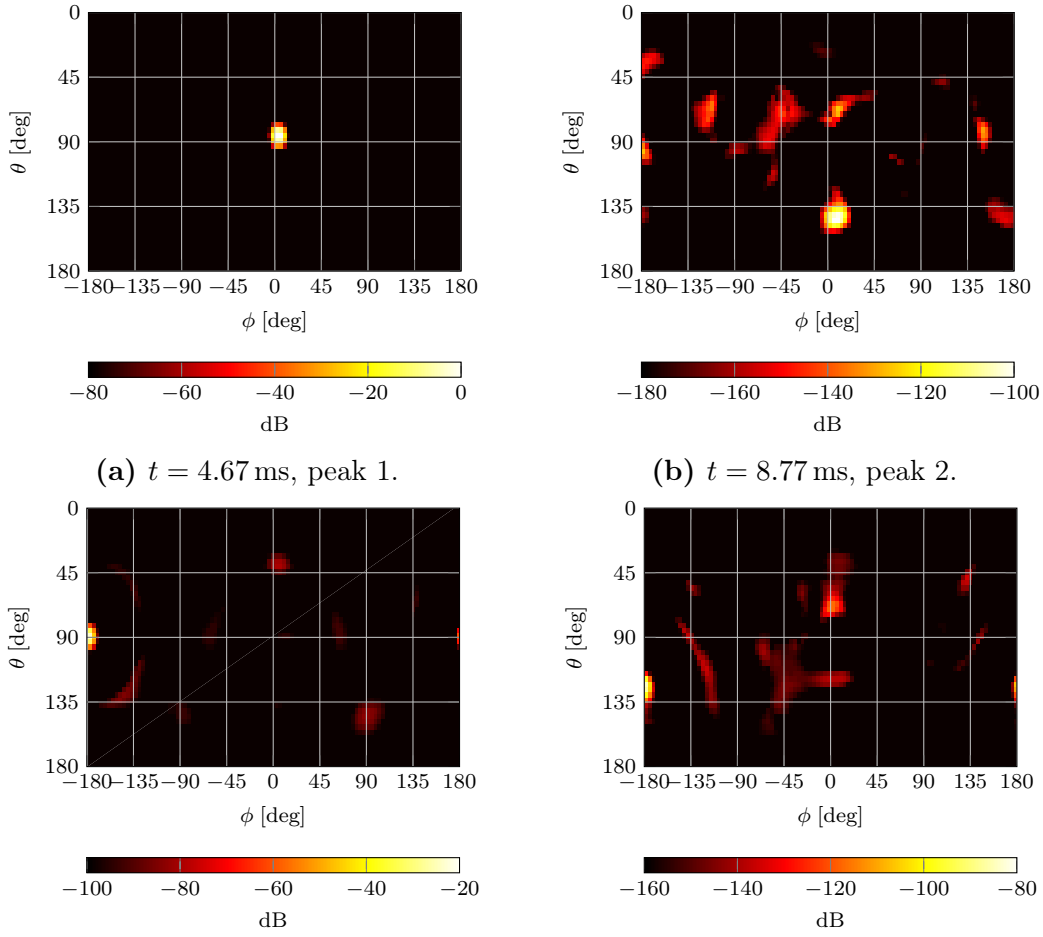
Finally, the power estimates in $\check{\mathbf{\Sigma}}$ are mapped onto a 2-D image using the equiangular projection [Snyder, 1993] (i.e. the azimuth angle is uniformly mapped to the horizontal axis while the polar angle is mapped to the vertical axis of a 2-D plot).

5.3.2 Validation in a Controlled Environment

The first set of experiments is conducted in an acoustically controlled environment (with reverberation time $T_{60} = 50\,\text{ms}$), according to the geometry depicted in Fig. 5.3a. The acoustic scene consists of a sound source placed at an height of 1.3 m, the spherical microphone array placed at the same height and distant 1.6 m from the sound source and a reflective panel placed 1 m behind the array. Reflections from the ceiling and from one wall (the one behind the sound source) have been damped through the use of absorbing panels. Reflections from the floor have been damped by a thick carpet.

Figure 5.3b shows the impulse response recorded by the first capsule of the microphone array. The peaks are labeled according to the acoustic paths shown in Fig. 5.3a. The location of the peaks in Fig. 5.3b is consistent with the length of the acoustic paths in Fig. 5.3a; in particular $\text{length}_1 = 1.6\,\text{m}$ ($\sim 4.7\,\text{ms}$), $\text{length}_2 = 3.05\,\text{m}$ ($\sim 8.8\,\text{ms}$), $\text{length}_3 = 3.6\,\text{m}$ ($\sim 10.5\,\text{ms}$), $\text{length}_4 = 4.4\,\text{m}$ ($\sim 12.8\,\text{ms}$). Acoustic paths reflected by the ceiling and the rear wall are not shown because they are so damped that their associated peaks are not visible in the impulse response.

Fig. 5.4 shows four panoramic views of acoustic reflections computed from different time segments, each centered around one of the peaks identified from Fig. 5.3b and associated to the acoustic path depicted in 5.3a. Fig. 5.4a is captured around $t = 4.67\,\text{ms}$, corresponding to the propagation time along the direct path 1 in Fig. 5.3a; a sharp peak in the estimated energy distribution is observed, coming from the direction under which the sound source is seen by the microphone array, i.e. $\theta = 90°, \phi = 0°$. Fig. 5.4b is computed from the time segment centered at $t = 8.77\,\text{ms}$, corresponding to the propagation time associated with the acoustic path 2, originated at the sound source and reflected by the floor the measurement chamber; this reflection is highly damped (due to the presence of a carpet) but still detectable with the proposed technique. Fig. 5.4c is captured around 10.7 ms, corresponding to the propagation time associated with path 3, i.e. the



(a) $t = 4.67\,\text{ms}$, peak 1. **(b)** $t = 8.77\,\text{ms}$, peak 2.

(c) $t = 10.70\,\text{ms}$, peak 3. **(d)** $t = 13.06\,\text{ms}$, peak 4.

Figure 5.4: Panoramic views of acoustic reflections in the acoustically controlled environment. The values of the power estimates expressed in dB are mapped to color scale.

first-order path passing through the reflective panel. Finally, Fig. 5.4d is captured around $t = 13.06\,\text{ms}$, corresponding to the propagation time associated with path 4, a second order path passing through the floor and the panel.

5.3.3 Experiments in a Real-World Acoustic Environment

The last set of experiments is conducted in the *Auditorium "Giovanni Arvedi"*, located in the *Museo del Violino* (Violin Museum), Cremona, Italy[1], whose floor plane is shown in Fig. 5.5. Dimensions of the hall are $14\,\text{m} \times 35\,\text{m}$ for a maximum height of 14 m, the volume of the hall being $5300\,\text{m}^3$. The sound source is placed in the center of the stage area (in the position marked by the cross) on a support of height 1.2 m, while the microphone array is placed in correspondence of the black dot in Fig. 5.5. The distance between the source and the center of the microphone array is 10.4 m.

The results of the application of the proposed analysis methodology to impulse

[1]We thank Museo del Violino Fondazione Stradivari and its staff for their availability during the measurement session held in the Auditorium G. Arvedi.

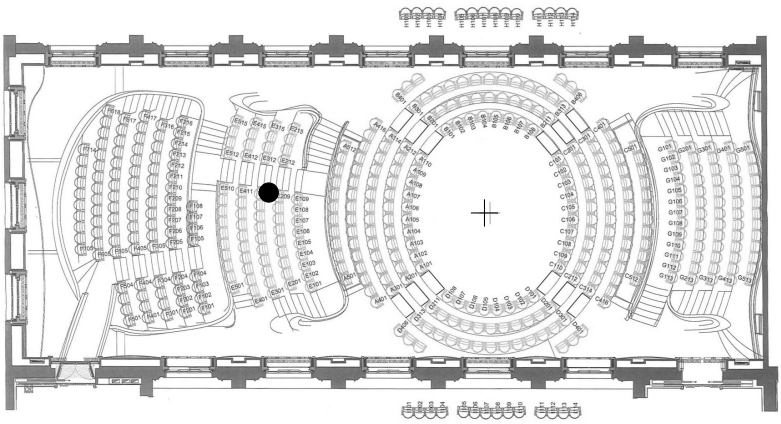
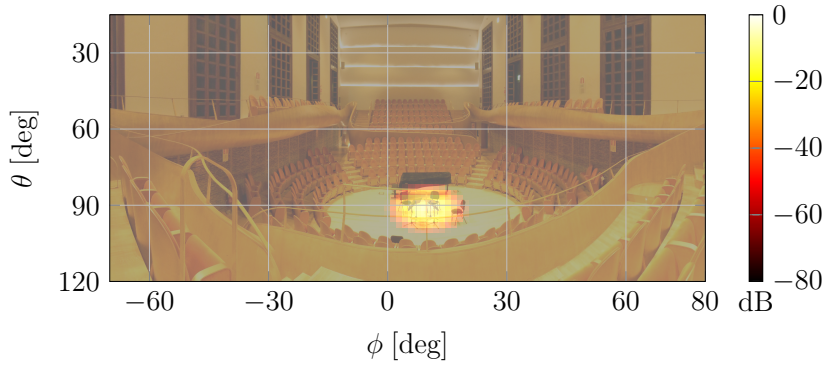
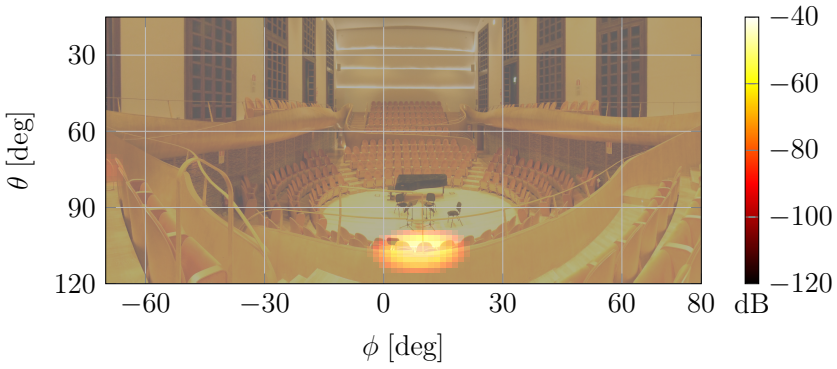


Figure 5.5: Floor plan of the auditorium "Giovanni Arvedi", *Museo del Violino*, Cremona, Italy.

responses acquired with the setup described above are shown in Fig. 5.6. The two acoustic



(a) $t = 30.32\,\text{ms}$, direct path.



(b) $t = 33.59\,\text{ms}$, railing diffraction.

Figure 5.6: Panoramic views of acoustic reflections in the auditorium "Giovanni Arvedi", *Museo del Violino*, Cremona, Italy. The values of the power estimates expressed in dB are mapped to color scale.

images are computed from two different time segments; the first one, Fig.5.6a, centered in $t = 30.32\,\text{ms}$, corresponds to the direct acoustic path from the sound source to the center of the microphone array; the other, Fig. 5.6b, centered in $t = 33.59\,\text{ms}$, corresponds to the

propagation time along the acoustic path generated from the sound source and diffracted by the railing. The acquired impulse responses do not exhibit other relevant events from the frontal direction, except the direct and diffraction paths. This is due to the design of the environment, which does not contain planar surfaces between the source and the seats, thus enabling only diffusive paths.

From the analysis of the above presented results, one can observe that the proposed methodology is able to discriminate acoustic events occurring within $\sim 3\,\text{ms}$, and coming from close directions ($\sim \deg 15$). These results demonstrate the direct applicability of the proposed approach to a challenging real-world scenario, in which the acoustic properties of the environment are not controlled.

5.4 Main results

- The analysis problem is formulated with reference to the plane wave representations discussed in Sec. 3.3.1. The integral formulation is discretized and adapted to accommodate both a short-time analysis framework and real-world noise added at microphone capsules to the recorded acoustic field.

- A solution to the analysis problem is presented based on a spherical microphone array. The technique is based on a state-of-the-art spectral analysis technique. The adoption of such technique, together with the short-time plane wave model enables highly accurate identification of acoustic reflections. An experimental validation is shown both in an acoustically conditioned and unconditioned environments.

Chapter 6

Synthesis of Acoustic Fields

The synthesis of acoustic fields is a goal that has been pursued by the signal processing community for decades. Solutions based on spatial distributions of loudspeakers constitute today a well-established technology, aimed at eliciting in the listener a deep sense of immersion [Spors et al., 2013]. In broad terms, what such solutions do is to reproduce a desired acoustic field through the superposition of the contributions of secondary sources, whose driving signals can be obtained analytically, as in Poletti [2005], Ahrens and Spors [2008a], Wu and Abhayapala [2009], Gupta and Abhayapala [2011], Berkhout et al. [1993], Spors et al. [2008], Ahrens and Spors [2010] or as a numerical approximation of the solution of an inverse problem, as in Kirkeby and Nelson [1993], Kirkeby et al. [1996], Ward and Abhayapala [2001], Kolundzija et al. [2009a], Antonacci et al. [2009], Lilis et al. [2010].

Among the analytical approaches, the family based on the spherical wave representation of the acoustic field is referred to as HOA [Poletti, 2005, Ahrens and Spors, 2008a, Wu and Abhayapala, 2009, Gupta and Abhayapala, 2011] and it is derived for circular and spherical loudspeaker deployments. Those falling into the category of WFS methods [Berkhout et al., 1993, Spors et al., 2008] rely on the Kirchhoff-Helmholtz integral equation (3.72). The original formulation of WFS, as found in Berkhout et al. [1993], was derived for linear loudspeaker deployments.

HOA and WFS approaches are optimal in terms of reproduction accuracy for the specific geometry that they were formulated for. Their extension to arbitrary geometries would either require a complete re-purposing of the approach (see Ahrens and Spors [2010] in the case of HOA for linear and planar loudspeaker arrays); or accept the presence of artifacts in the rendered acoustic field (see Spors and Rabenstein [2006], Spors et al. [2008], Ahrens and Spors [2012] for WFS). Besides approaches based on HOA and WFS, Hannemann and Donohue [2008] proposes an approach for the rendering of acoustic fields in a small listening region centered on the listener's head with an arbitrary distribution of loudspeakers.

The other class of approaches, instead of adopting analytic representations, is based on the numerical approximation of the desired acoustic field on a set of control points, e.g. with least squares [Kirkeby and Nelson, 1993, Kirkeby et al., 1996, Ward and Abhayapala, 2001] or sparse linear regression [Lilis et al., 2010] techniques. Most of these approaches require the solution of an inverse problem that is not guaranteed to be well-conditioned, as pointed out in Fazi and Nelson [2007], Kolundzija et al. [2009a]; for this reason, regularization techniques are usually employed to find a stable solution, as in Poletti [2005], Kolundzija et al. [2009b], Antonacci et al. [2009]. The impact of regularization parameters on physical and perceptual properties of the reproduced acoustic field is a

subject of ongoing research [Spors et al., 2013]. This family of approaches presents the advantage of not being constrained to a specific geometry of the reproduction setup.

This chapter proposes an analytic solution to the problem of acoustic field rendering, which leverages on specific geometrical aspects of the underlying problem. This approach is based on the plane wave representation of the acoustic field. Indeed, loudspeakers driving functions are derived from the plane wave components through simple geometrical considerations. Relying on simple and intuitive geometrical considerations show that the definition of driving functions for arbitrary arrangements of loudspeakers along a planar curve is possible. More specifically, this chapter provides closed-form expressions for the plane wave representation of two commonly used acoustic field models (plane-wave acoustic fields and acoustic fields due to a point source), parametrized by the source location. We interpret the plane wave representation of Whittaker-type as a function describing amplitude and phase of individual plane-wave components that contribute to the desired acoustic field. We then exploit the overall loudspeaker array to reproduce individual plane-wave components, properly weighed by the plane wave expansion coefficients. The desired acoustic field is then obtained as the superposition of these reproduced components.

This chapter focuses on finding the proper driving functions to reproduce 2-D acoustic fields. In this scenario, we show the relationship between the number of loudspeakers, the desired accuracy, the area of the reproduction region and the temporal frequency range, following the analysis presented in Ward and Abhayapala [2001], Kennedy et al. [2007]. The concepts presented here can be extended to the 3-D case by simply considering a 3-D plane wave representation and suitable plane-wave rendering methods for 3-D deployments of loudspeakers. Although in this chapter only a free-field propagation scenario is considered, i.e. a scenario in which the effect of reverberation is ignored, the presented approach can readily be integrated with state-of-the-art compensation techniques to attenuate the impact of room reverberations on the rendering accuracy, e.g. using methods in Carini et al. [2012], Canclini et al. [2014a,b], Poletti et al. [2015].

The proposed rendering technique will, at first, be presented in an ideal setting, where loudspeakers are assumed to be unrealistic height-invariant sources (i.e. sources propagating according to 3.61) arranged in space with infinite accuracy. However, in any practical situation none of the above assumptions is realistic and strategies must be devised in order to control the impairments introduced by such modeling mismatches. For this purpose, a technique is proposed that exploits the non idealities of the loudspeaker array in order to provide a more realistic solution. In particular, the attention is focused on the relevant case of a loudspeaker propagation function that is different from the ideal one and, possibly, it varies among the loudspeakers. Analysis of the data provided by a loudspeaker manufacturer show that, indeed, the propagation function of a set of loudspeakers can be conveniently modeled as a random variable, being the propagation of each loudspeaker subject to many non-idealities that may appear in the manufacturing process. The presented technique relies on the statistics of the propagation functions in order to derive a stable solution for the loudspeaker driving functions.

The work presented in this chapter has been published by the authors in Bianchi et al. [2014, 2016].

6.1 Preliminaries and Problem Statement

This section discusses how to use the plane wave representation of an acoustic field for generating a desired acoustic field in a given region. This is done for the 2-D case, meaning that the field is assumed to be invariant along the vertical direction, as in Ahrens and

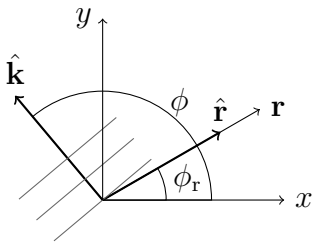


Figure 6.1: Position vector $\mathbf{r}$ and wave vector $\hat{\mathbf{k}}$.

Spors [2008a], Wu and Abhayapala [2009].

6.1.1 Notation

This paragraph briefly reviews the notation adopted in the rest of the thesis and here specialized to the problem under consideration. With reference to Fig. 6.1, let $\mathbf{r} = [r\cos(\phi_{\mathrm{r}}), r\sin(\phi_{\mathrm{r}})]^T$ denote the position vector in the 2-D space $\mathbb{R}^2$. The unit vector in the direction of the (nonzero) vector $\mathbf{r}$ is defined as $\hat{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|$, where the notation $\|\cdot\|$ is to be interpreted as $\|\cdot\|_2$.

6.1.2 Plane-Wave Model

Following Sec. 4.1.1, an arbitrary acoustic field can be modeled as the integral of plane waves. In an height-invariant scenario, it follows that $k_z = 0$, hence the wavevector can be parametrized as $\mathbf{k}(\phi) = (\omega/c)[\cos(\phi), \sin(\phi)]$, $\phi \in [0, 2\pi)$. Starting from (4.3) it is straightforward to verify that, in two-dimensions, an arbitrary acoustic field can be modeled as the integral of plane waves coming from directions $\mathbf{k}(\phi)$, $\phi \in [0, 2\pi)$

$$P(\mathbf{r}, \omega) = \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} e^{j\frac{\omega}{c}\langle \mathbf{r}, \hat{\mathbf{k}}(\phi)\rangle} \tilde{P}(\phi, \omega)\, d\phi, \tag{6.1}$$

where $\tilde{P}(\phi, \omega)$ is a complex-valued function that modulates each plane wave component in amplitude and phase. As it as been reviewed in Sec. 4.1.1, this function is usually referred to as *Herglotz density* [Colton and Kress, 1992, p. 54].

In this chapter the acoustic field is described through its Herglotz density $\tilde{P}(\phi, \omega)$. Indeed, $\tilde{P}(\phi, \omega)$ is space invariant in a region that is free of scatterers (cfr. 4.1.1). In Sec. 6.3 the spatial invariance of $\tilde{P}(\phi, \omega)$ is exploited in order to compute the loudspeakers driving signals.

6.1.3 Problem Statement

The main goal of this chapter is to show how to use the plane-wave representation for rendering, in a region of interest $\mathcal{S} \subset \mathcal{V}$, the acoustic field produced by a source placed outside $\mathcal{S}$ (primary source). In particular, the attention is focused on finding an expression for $\tilde{P}(\mathbf{r}, \omega)$ given the parameters of a desired acoustic source. Parameters considered here are the spatial position of the source and its type (point source or plane wave). The loudspeakers (secondary sources) that generate the acoustic field are assumed to be placed on the border $\partial\mathcal{S}$ of the region of interest.

In order to address the problem with a certain progression, first the theoretical analysis and then its implementation are discussed. At first, it is assumed that ideal plane-wave

generators are available. Then, it is shown how to implement such ideal generators using a spatial distribution of loudspeakers in Section 6.3.2. We also discuss the errors introduced with the discretization of the integral (6.1), and approximating ideal plane wave generators with spatial arrangements of loudspeakers.

6.2 Model-Based Analysis of the Herglotz density function

This section derives closed-form expressions for $\tilde{P}(\phi,\omega)$ in the cases of a source at infinity (plane wave) and an (isotropic) point source at finite distance. These two cases are important because complex acoustic scenes are, indeed, combinations of such sources. In fact, sources with non-isotropic radiation pattern can be treated as compact spatial distributions of point sources [Ahrens and Spors, 2007a,b, Corteel, 2007]. Furthermore, complex acoustic scenes can always be constructed as combinations of (generally anisotropic) real and image sources, which can be accounted for using the superposition principle.

For the sake of clarity, in this sections all the parameters related to the desired virtual source (i.e. source position or DoA)are denoted by the subscript $(\cdot)_\mathrm{z}$.

In general, the density $\tilde{P}(\phi,\omega)$ is periodic (of period 2π) in ϕ, therefore it can be Fourier-expanded as

$$\tilde{P}(\phi,\omega) = \sum_{m=-\infty}^{\infty} A_m(\omega) e^{jm\phi}, \tag{6.2}$$

where $A_m(\omega)$ are the Fourier series coefficients. An arbitrary acoustic field can be completely reconstructed from the sole knowledge of the coefficients $A_m(\omega)$, cfr. (4.45)

$$P(\mathbf{r},\omega) = \sum_{m=-\infty}^{\infty} J_m\left(\frac{\omega}{c}\rho\right) A_m(\omega) e^{jm\phi_\mathrm{r}}, \tag{6.3}$$

$J_m(\cdot)$ being the Bessel function of first kind and order m [Abramowitz and Stegun, 1972, Chap. 9] and $\rho = \sqrt{x^2+y^2}$. In the following it is shown that, given the model of a desired acoustic field (planar source or point source), the Herglotz density function can be computed through (6.2).

6.2.1 Plane Wave

Consider a plane wave coming from direction $\hat{\mathbf{z}} = [\cos(\phi_\mathrm{z}), \sin(\phi_\mathrm{z})]^T$. We assume that $\hat{\mathbf{z}}$ has no vertical component ($\mathcal{S}$ is assumed as lying on the horizontal plane). The acoustic field in $\mathbf{r} \in \mathcal{S}$ is [Williams, 1999, Eq. (2.24)], (3.16)

$$P(\mathbf{r},\omega) = e^{j\frac{\omega}{c}\langle \mathbf{r},\hat{\mathbf{z}}\rangle} = \sum_{m=-\infty}^{\infty} J_m\left(\frac{\omega}{c}\rho\right) e^{jm(\phi_\mathrm{r}-\phi_\mathrm{z})}, \tag{6.4}$$

where the second equality is obtained thanks to the Jacobi-Anger expansion [Abramowitz and Stegun, 1972, Eq. (9.1.41)]. The coefficients $A_m(\omega)$, in this case, are obtained by matching mode by mode (6.4) and (6.3), obtaining

$$A_m(\omega) = e^{-jm\phi_\mathrm{z}}. \tag{6.5}$$

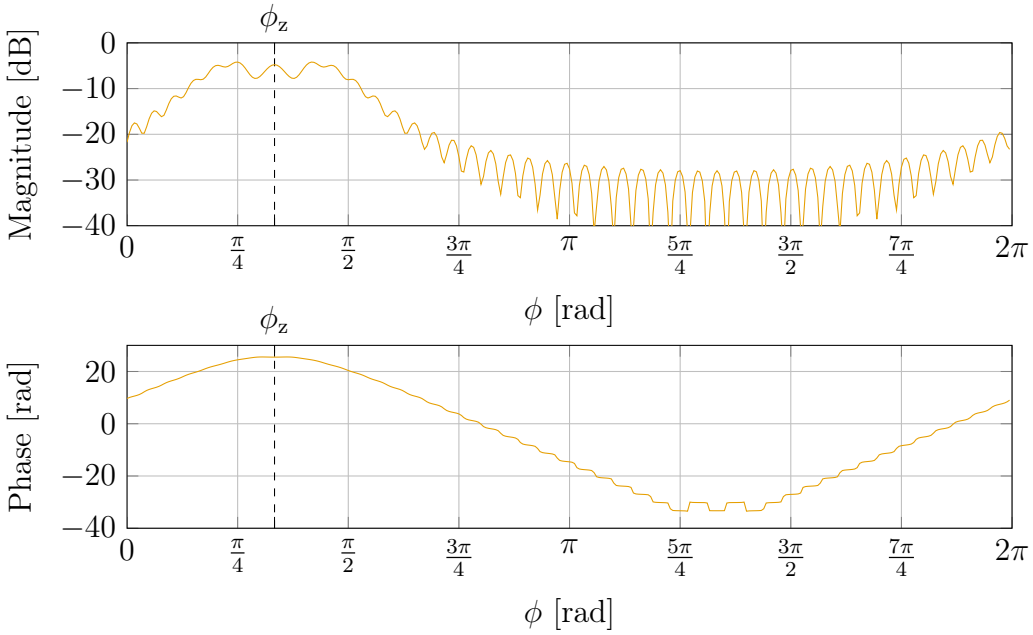


Figure 6.2: Magnitude and phase of the Herglotz density associated to a point source in $\mathbf{r}_z = [1, \sqrt{3}]^T$. Gibbs phenomena are due to the fact that the Series expansion in (6.2) has been truncated to order $M = 25$, $(m = -25, \ldots, 25)$.

6.2.2 Point Source

Consider now an isotropic 2-D point source (a source that produces a vertically invariant acoustic field). The acoustic field due to such point source in $\mathbf{r}_z = \rho_z[\cos(\phi_z), \sin(\phi_z)]^T$ is given by [Williams, 1999, Chap. 4], (3.61)

$$P(\mathbf{r}, \omega) = \frac{j}{4} H_0^{(1)} \left(\frac{\omega}{c} \|\mathbf{r} - \mathbf{r}_z\| \right), \tag{6.6}$$

The Hankel function in (6.6) can be expanded using the addition theorem [Abramowitz and Stegun, 1972, Eq. (9.1.75)], giving

$$P(\mathbf{r}, \omega) = \sum_{m=-\infty}^{\infty} \frac{j}{4} J_m \left(\frac{\omega}{c} \rho \right) H_m^{(1)} \left(\frac{\omega}{c} \rho_z \right) e^{jm(\phi_r - \phi_z)}, \tag{6.7}$$

where $H_m^{(1)}(\cdot)$ is the Hankel function of first kind and order m. The coefficients $A_m(\omega)$ are obtained by matching mode by mode (6.7) and (6.3) as

$$A_m(\omega) = \frac{j}{4} H_m^{(1)} \left(\frac{\omega}{c} r_z \right) e^{-jm\phi_z}. \tag{6.8}$$

Fig. 6.2 shows the Herglotz density associated to a point source in $\mathbf{r}_z = [1, \sqrt{3}]^T$, with parameters $\omega = 2\pi f$, $f = 1\,\text{kHz}$, $A(\omega) = 1$. The Herglotz density has been computed by substituting (6.8) into (6.2). The infinite series in (6.2) has been truncated to $M = 25$ $(m = -25, \ldots, 25)$, which introduces Gibbs phenomena whose impact on rendering will be discussed in Sec. 6.3.

The expressions of $\tilde{P}(\phi, \omega)$ obtained for plane waves and point sources are used in the next sections in order to derive the rendering filters.

6.3 Acoustic Rendering with a Circular Array

This section describes how an arbitrary acoustic field can be reproduced using directional acoustic field modeling. Indeed, the tools presented in Sec. 6.2 provide a parametric model for the acoustic scene to be rendered. In particular, (6.4) or (6.8) enable to compute the plane-wave density function based on the model of the desired acoustic field.

6.3.1 Discrete Distribution of Plane Waves

Consider a continuous distribution of plane waves propagating from directions $\mathbf{s}(\phi) = [\rho\cos(\phi), \rho\sin(\phi)]^T$. The decomposition of the acoustic field in terms of this set of plane waves is given by (6.1), which states that the acoustic field at any location in $\mathcal{S}$ is obtained by simply knowing the Herglotz density of the target acoustic field valid on the scatterer-free region $\mathcal{S}$. This is made possible through the spatial invariance of $\tilde{P}(\phi,\omega)$, which allows us to reconstruct the acoustic field without any additional information other than the knowledge of $\tilde{P}(\phi,\omega)$.

Consider now a discrete distribution of plane waves coming from directions $\mathbf{s}(\phi_n)$, $n = 1,\ldots,N$ that uniformly samples the interval $[0, 2\pi)$. In this case the acoustic field in (6.1) is approximated by

$$P_N(\mathbf{r},\omega) = \frac{1}{N}\sum_{n=1}^{N} \tilde{P}(\phi_n,\omega) e^{j\frac{\omega}{c}\langle \mathbf{r},\hat{\mathbf{k}}(\phi_n)\rangle}. \tag{6.9}$$

In order to understand the properties of the discrete expansion in (6.9), in particular to quantify the discretization error, one reverts to the modal representation introduced in Sec. 6.2. Notice that $\tilde{P}(\phi,\omega)$, as given by (6.2), is periodic in ϕ with period 2π. It is common to approximate the infinite series in (6.2) through a truncation of the modal expansion to order M (i.e. $m = -M,\ldots,M$). In this situation, the highest angular frequency contributing to $\tilde{P}(\phi,\omega)$ is $M/(2\pi)$, which means that the highest-frequency mode is $e^{jM\phi}$ [Ward and Abhayapala, 2001, Kennedy et al., 2007, Wu and Abhayapala, 2009]. As previously recalled in 4.2.3, in Kennedy et al. [2007, Eq. 24] it is proved that

$$|P(\mathbf{r},\omega) - P_N(\mathbf{r},\omega)| < 0.16127 \int_0^{2\pi} |\tilde{P}(\phi,\omega)|\, d\phi \quad \text{if} \quad M \geq \left\lceil e\frac{\omega}{c}\frac{\rho}{2} \right\rceil. \tag{6.10}$$

Eq. (6.10) enables to bound the normalized truncation error given the radius ρ of the listening area and the frequency ω. According to the Shannon sampling theorem, the acoustic field can be exactly reconstructed from its samples if the angular interval between adjacent plane-wave directions of arrival satisfies $\phi_n - \phi_{n-1} < \pi/M\,, n = 2,\ldots,N$. For the case of N uniformly distributed plane waves, if $N \geq 2M+1$ the acoustic field can be reconstructed without introducing any further errors other than the one introduced by the discretization.

The approximation error due to discretization can be computed as the mean square error between the desired acoustic field $P(\mathbf{r},\omega)$ and the discretized one $P_N(\mathbf{r},\omega)$

$$\varepsilon_N(\rho,\omega) = \frac{1}{\pi\rho^2}\int_0^{\rho}\int_0^{2\pi} \frac{|P(\mathbf{r},\omega) - P_N(\mathbf{r},\omega)|^2}{|P(\mathbf{r},\omega)|^2}\, d\phi_{\mathrm{r}}\, d\rho. \tag{6.11}$$

Figure 6.3 shows the approximation error (6.11) as a function of $\frac{\rho}{c}r$ for different numbers N of plane-wave components, all satisfying the condition $N = 2M+1$ for $M = 0, 5, 10, 15, 20, 25$. The target acoustic field is that produced by a 2-D point source in

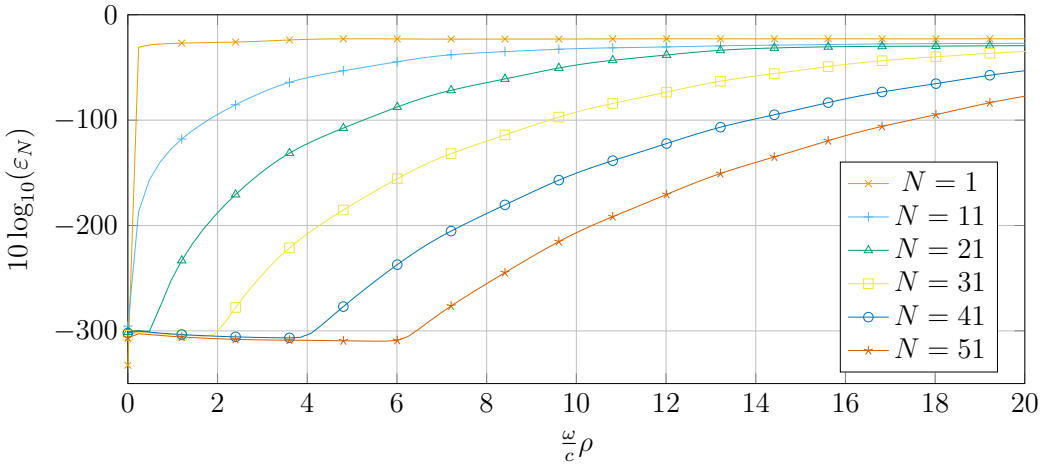


Figure 6.3: Mean square discretization error (6.11). N denotes the number of plane-wave components. For each curve the truncation order is set to $M = \lfloor (N-1)/2 \rfloor$.

$\mathbf{r}_z = [1, \sqrt{3}]^T$ (i.e. $\rho_z = 2\,\mathrm{m}$ and $\phi_z = \pi/3$). The speed of sound is set to $c = 340\,\mathrm{m\,s^{-1}}$. As expected, the larger is the reproduction area, the larger the number N of secondary sources used to keep ε_N down to reasonable levels throughout the whole listening area.

6.3.2 Implementation with Circular Arrays

This paragraph focuses on the reproduction of a plane wave with direction of arrival ϕ_n with an array of L 2-D loudspeakers (i.e. transducers ideally capable of generating an height invariant acoustic field) that sample the circular contour $\partial\mathcal{S}$ of the listening area. The same L loudspeakers are used for reproducing each one of the N plane waves at the same time, using the superposition principle. This enables to save in the number of transducers, while guaranteeing that the maximum number of loudspeakers is used for approximating each secondary source.

In order to generate a plane wave coming from direction ϕ_n, the array of loudspeakers must be driven by an appropriate spatial filter. Among all possible choices, here we adopt the one presented in Wu and Abhayapala [2009], which has been proven to provide good performance for circular rendering setups. However, any other spatial filtering technique capable of generating plane waves from a circular array of loudspeakers (e.g. the one presented in Ward and Abhayapala [2001]) could be used with no relevant changes in the rendering approach. The derivation of the spatial filter follows the mode-matching strategy already adopted in Sec. 6.2.

Assume that loudspeakers are placed at positions $\mathbf{v}_l = \rho[\cos(\phi_l), \sin(\phi_l)]$, $l = 1, \ldots, L$, thus uniformly sampling the contour $\partial\mathcal{S}$. The filter of the lth loudspeaker is given by

$$h_l(\phi_n, \omega) = \frac{4}{jL} \sum_{m=-M}^{M} \frac{e^{jm(\phi_l - \phi_n)}}{H_m^{(1)}\left(\frac{\omega}{c}\rho\right)}. \tag{6.12}$$

As shown in Wu and Abhayapala [2009], $\mathbf{h}(\phi_n, \omega) = [h_1(\phi_n, \omega), \ldots, h_L(\phi_n, \omega)]^T$ enables to approximate a plane wave of direction ϕ_n in $\mathcal{S}$, as

$$e^{j\frac{\omega}{c}\langle \mathbf{r}, \hat{\mathbf{k}}(\phi_n)\rangle} \approx \sum_{l=1}^{L} h_l(\phi_n, \omega) G(\mathbf{r}|\mathbf{v}_l, \omega), \quad \mathbf{r} \in \mathcal{S}, \tag{6.13}$$

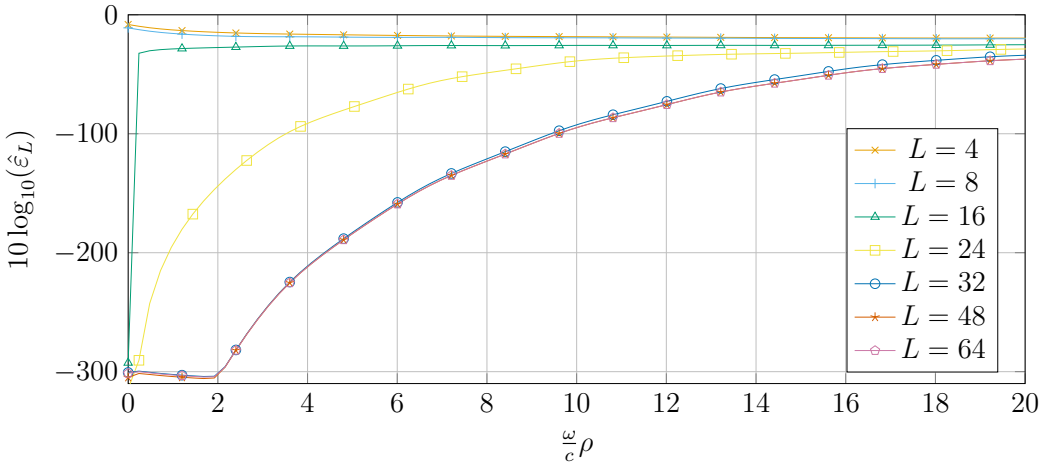


Figure 6.4: Mean square approximation error (6.16) as a function of $\frac{\omega}{c}\rho$. L is the number of loudspeakers.

$G(\mathbf{r}|\mathbf{v}_l, \omega)$ being the Green's function from the loudspeaker in $\mathbf{v}_l$ to the point $\mathbf{r}$ (cfr. (3.61))

$$G(\mathbf{r}|\mathbf{v}_l, \omega) = \frac{j}{4} H_0^{(1)} \left(\frac{\omega}{c} \|\mathbf{r} - \mathbf{v}_l\| \right). \tag{6.14}$$

By substituting (6.13) into (6.9) one obtains that the desired acoustic field at position $\mathbf{r}$ is approximated by

$$\hat{P}_L(\mathbf{r}, \omega) = \frac{1}{N} \sum_{n=1}^{N} \tilde{P}(\phi_n, \omega) \sum_{l=1}^{L} h_l(\phi_n, \omega) G(\mathbf{r}|\mathbf{v}_l, \omega). \tag{6.15}$$

In order to analyze quantitatively the performance of the spatial filter (6.12), the approximation error between the desired acoustic field $P(\mathbf{r}, \omega)$ and the approximated one $\hat{P}_L(\mathbf{r}, \omega)$ is introduced as

$$\hat{\varepsilon}_L(\rho, \omega) = \frac{1}{\pi r^2} \int_0^{\rho} \int_0^{2\pi} \frac{|P(\mathbf{r}, \omega) - \hat{P}_L(\mathbf{r}, \omega)|^2}{|P(\mathbf{r}, \omega)|^2} \, d\phi_{\mathrm{r}} \, d\rho. \tag{6.16}$$

Fig. 6.4 shows this error for different numbers of loudspeakers L. In these simulations the number of modes and the number of plane waves is fixed ($M = 15$ and $N = 32$) while the number of loudspeakers is let vary between 4 and 64. In all cases, the loudspeakers uniformly sample a circle of radius $\rho = 1\,\mathrm{m}$. The target acoustic field is that of a 2-D point source in $\mathbf{r}_{\mathrm{z}} = [1, \sqrt{3}]^T$ (i.e. $\rho_{\mathrm{z}} = 2\,\mathrm{m}$ and $\phi_{\mathrm{z}} = \pi/3$) at a frequency $\omega = 2\pi f$, $f = 1\,\mathrm{kHz}$. The speed of sound is set to $c = 340\,\mathrm{m\,s^{-1}}$.

Fig. 6.4 shows that the best choice for the number of loudspeakers is $L = N = \lceil 2M+1 \rceil$, while using $L > N = 32$ loudspeakers does not introduce further significant error reductions. As mentioned above, these L loudspeakers are all used for rendering all the N plane waves at once. From (6.15), in fact, we obtain that the driving signal for the lth loudspeaker that render all the N plane-wave components is

$$w_l(\omega) = \frac{1}{N} \sum_{n=1}^{N} \tilde{P}(\phi_n, \omega) h_l(\phi_n, \omega). \tag{6.17}$$

Fig. 6.5 shows the block diagram for the computation of the overall driving signals. Notice that the filters $h_l(\phi_n, \omega)$, are completely independent of the sound scene to be rendered, and can therefore be pre-computed. What changes are only the weights $\tilde{P}(\phi_n, \omega)$, which carry the information on the actual acoustic scene to be rendered.

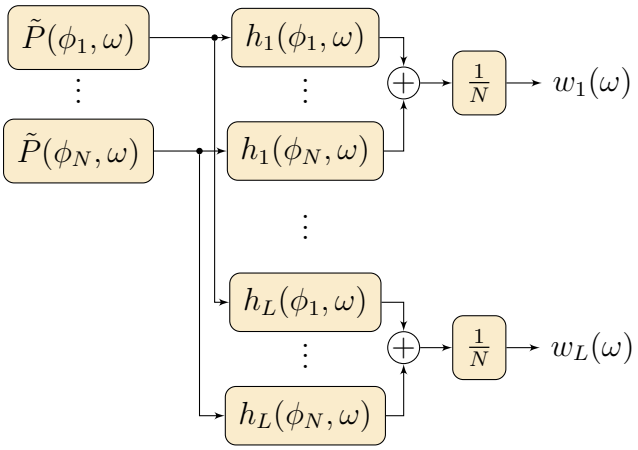


Figure 6.5: Implementation of the loudspeaker filters for the simultaneous rendering of all secondary sources. The set of density coefficients $\{\tilde{P}(\phi_n, \omega)\}$ is independent on the geometry of the rendering system, while the set of spatial filters $\{h_l(\phi_n, \omega)\}$ is independent on the actual acoustic field to be reproduced.

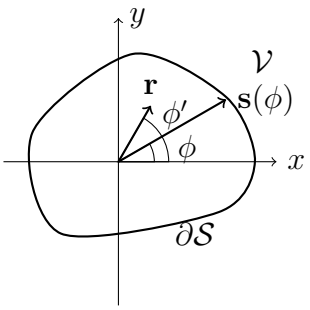


Figure 6.6: Secondary sources on a convex curve.

6.4 Generalization to Non-Circular Arrays

In order to demonstrate the generalization of the proposed approach to non-circular geometries of the loudspeaker array, this section discusses the realization of a rendering system deployed on two illustrative geometries: a generic convex curve and a line segment.

6.4.1 Convex Curve

Consider a region of interest $\mathcal{S}$ that is a convex bounded subset of $\mathcal{V}$, as depicted in Fig. 6.6. In this geometry, the entire discussion presented for the case of a circular contour in Sec. 6.3 holds with slight adaptations. In fact, if loudspeakers are placed on a closed convex curve, they can still be used to effectively reproduce the set of plane waves that uniformly samples the interval $\phi \in [0, 2\pi)$. Indeed, the spatial invariance of the Herglotz density function enables to conceive points on the border of a convex curve as unique plane-wave directions of arrival. Each plane wave is modulated by $\tilde{P}(\phi, \omega)$ as in the case of the circular rendering setup. The only difference with respect to the circular case lies in the computation of the spatial filter, as we will show in Paragraph 6.4.3.

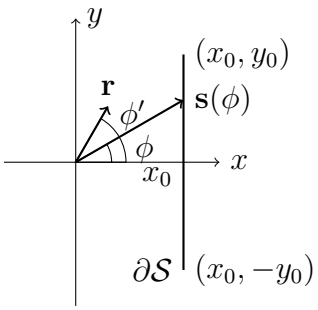


Figure 6.7: Secondary sources on a line segment.

6.4.2 Line Segment

Consider now the distribution of secondary sources on the line segment $\mathbf{s} = [x_0, -y_0 \leq y \leq y_0]^T$, as depicted in Fig. 6.7. In this case the contour $\partial\mathcal{S}$ can still be parametrized by ϕ, but the allowed values for ϕ belong to a subset of $[0, 2\pi)$. In particular, from geometrical considerations it follows that the allowed range for ϕ is

$$\phi \in \{\mathbb{R} | \phi_{\min} \leq \phi \leq \phi_{\max}\}, \tag{6.18}$$

being $\phi_{\min} = \arctan(-y_0/x_0)$ and $\phi_{\max} = \arctan(y_0/x_0)$. The acoustic field due to the limited set of directions of propagation is

$$P(\mathbf{r}, \omega) = \frac{1}{2\pi} \int_{\phi_{\min}}^{\phi_{\max}} \tilde{P}(\phi, \omega) e^{j\frac{\omega}{c}\langle \mathbf{r}(\phi), \hat{\mathbf{k}}(\phi)\rangle} d\phi. \tag{6.19}$$

We refer the interested reader to Devaney and Sherman [1973], Shono and Inuzuka [1978] for a discussion of the conditions under which an acoustic field can be expressed with a limited number of plane waves.

In analogy with what we have done for the circular case (Sec. 6.3.1) and for the convex curve (Sec. 6.4.1), ϕ is sampled in an equiangular fashion with N plane-wave components. The discretization yields

$$P_N(\mathbf{r}, \omega) = \frac{\phi_{\max} - \phi_{\min}}{2\pi N} \sum_{n=1}^{N} \tilde{P}(\phi_n, \omega) e^{j\frac{\omega}{c}\langle \mathbf{r}, \hat{\mathbf{k}}(\phi_n)\rangle}. \tag{6.20}$$

In order to realize such rendering system, a loudspeaker array with L loudspeakers uniformly spaced on the y axis is considered, i.e. $\mathbf{v}_l = -y_0 + 2y_0(l-1)/(L-1)$. As in Sec. 6.4.1, all the loudspeakers contribute to render plane waves in any direction ϕ_n. Also in this case any spatial filter $\mathbf{h}(\phi_n, \omega)$ suitable to render plane waves with linear loudspeaker arrays (e.g. one of the techniques proposed in Berkhout et al. [1993], Ahrens and Spors [2010], Kumatani et al. [2010], Lee et al. [2015]) can be used to generate plane waves with direction ϕ_n, thus providing an approximation of the propagation term in (6.20). In Sec. 6.7.2 a novel technique to generate plane waves in real world scenarios is presented. Finally, the acoustic field is approximated by

$$\hat{P}_L(\mathbf{r}, \omega) = \frac{\phi_{\max} - \phi_{\min}}{2\pi N} \sum_{n=1}^{N} \tilde{P}(\phi_n, \omega) \sum_{l=1}^{L} h_l(\phi_n, \omega) G(\mathbf{r} | \mathbf{v}_l, \omega). \tag{6.21}$$

6.4.3 Implementation

As discussed above, the acoustic field rendering approach presented in this manuscript is independent on the actual technique used to produce plane waves from loudspeaker arrays. In Sec. 6.3.2, while describing the implementation for a circular array of loudspeakers, a spatial filter with a closed-form solution available from the literature was adopted. In particular, the filter was given by (6.12) and was derived from Wu and Abhayapala [2009]. When considering loudspeaker distributions of different shapes, such as the ones discussed in Sec. 6.4.1 and in Sec. 6.4.2, closed-form solutions are not, at our knowledge, available from the literature. In order to provide the reader with a simple solution, in this paragraph we show a basic spatial filter design.

This paragraph discusses a naive technique to design the spatial filter $\mathbf{h}(\phi_n, \omega)$ in such a way to minimize the error due to the approximation of plane-wave acoustic fields with loudspeakers, i.e. by finding the spatial filter $\hat{\mathbf{h}}(\phi_n, \omega)$ that minimizes the approximation error in (6.13). In mathematical terms,

$$\hat{\mathbf{h}}(\phi_n, \omega) = \underset{\mathbf{h}(\phi_n,\omega)}{\arg\min} \left\| e^{j\frac{\omega}{c}\langle \mathbf{r}, \hat{\mathbf{k}}(\phi_n)\rangle} - \sum_{l=1}^{L} h_l(\phi_n, \omega) G(\mathbf{r}|\mathbf{v}_l, \omega) \right\|^2 . \tag{6.22}$$

The optimization problem in (6.22) can be solved with a least-squares approach, whose solution is analytically given by writing the problem in matrix form, with $\{\mathbf{d}(\phi_n, \omega)\}_i = e^{j\frac{\omega}{c}\langle \mathbf{r}_i, \hat{\mathbf{k}}(\phi_n)\rangle}$ denoting the desired plane-wave field over a set of control points $\mathbf{r}_i$, $i = 1, \ldots, I$; $\{\mathbf{G}(\omega)\}_{i,l} = G(\mathbf{r}_i|\mathbf{v}_l, \omega)$ being a matrix containing the Green's functions from a loudspeaker in $\mathbf{v}_l$ to a control point in $\mathbf{r}_i$. With this notation, the solution to the problem in (6.22) can be written as

$$\hat{\mathbf{h}}(\phi_n, \omega) = (\mathbf{G}(\omega)^H \mathbf{G}(\omega) + \lambda \mathbf{I})^{-1} \mathbf{d}(\omega), \tag{6.23}$$

where a Tikhonov regularization have been adopted for the computation of the pseudo-inverse of the matrix $\mathbf{G}(\omega)$.

Section 6.5 will make use of (6.12), for the circular case, and (6.23), for the general case, to obtain spatial filters suitable to a simulative validation of the overall rendering approach. Notice that, in both cases, the spatial filters $\hat{\mathbf{h}}(\phi_n, \omega)$, $n = 1, \ldots, N$ can be pre-computed, as they are independent on the actual scene to be reproduced.

6.5 Simulation in Ideal Conditions

This section provides simulative results to evaluate the accuracy of the rendering technique proposed in this chapter under ideal conditions. We consider the cases of circular, elliptical and linear deployments of speakers. In order to enable a fair comparison, the results shown here are obtained with the state-of-the-art spatial filters introduced in Sec. 6.3-6.4 and not with the spatial filter presented in Sec. 6.7.2. It is assumed that loudspeakers able to generate the acoustic field in (3.61) are available.

6.5.1 Acoustic Field Synthesis with a Circular Array

Consider a circular reproduction region $\mathcal{S}$ of radius $\rho = 1\,\text{m}$. The target sound pressure field is that of a point source in $\mathbf{r}_\text{z} = [1, \sqrt{3}]$, i.e. $r_\text{z} = 2\,\text{m}$ and $\phi_\text{z} = \pi/3$ (Fig. 6.8a). The L loudspeakers are equally spaced on the contour $\partial\mathcal{S}$. In this simulation we set $L = N = 32$ and the truncation order is set to $M = 15$. Notice that this setup satisfies the condition

$L = N > 2M + 1$ discussed in Section 6.3. The position of the loudspeakers corresponds to the direction of propagation of plane waves, i.e. $\mathbf{v}_l = \mathbf{s}(\phi_l)$.

The symbol $\hat{P}_{\text{PWD}}(\mathbf{r}, \omega)$ denotes the acoustic field synthesized through (6.15). In order to compare our solution with state-of-the-art rendering techniques, we also consider WFS [Spors et al., 2008, eq. (23)]

$$\hat{P}_{\text{WFS}}(\mathbf{r}, \omega) = \frac{1}{\eta} \sum_{l=1}^{L} G(\mathbf{r}|\mathbf{v}_l, \omega) \frac{1}{2c} \alpha(\mathbf{v}_l) \frac{\langle \mathbf{v}_l - \mathbf{z}, \hat{\mathbf{k}}(\phi_l) \rangle}{\|\mathbf{v}_l - \mathbf{z}\|} j \frac{\omega}{c} H_1^{(1)} \left(\frac{\omega}{c} \|\mathbf{v}_l - \mathbf{z}\| \right), \tag{6.24}$$

being

$$\alpha(\mathbf{v}_l) = \begin{cases} 1, & \text{if } \langle \mathbf{v}_l - \mathbf{z}, \hat{\mathbf{k}}(\phi_l) \rangle < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{6.25}$$

and $\eta = \sum_l \alpha(\mathbf{v}_l)$ the number of active loudspeakers; and by HOA rendering methods [Wu and Abhayapala, 2009, eq. (20)]

$$\hat{P}_{\text{HOA}}(\mathbf{r}, \omega) = \frac{2\pi}{L} \sum_{l=1}^{L} G(\mathbf{r}|\mathbf{v}_l, \omega) \sum_{m=-M}^{M} \frac{2}{j\pi H_m^{(1)} \left(\frac{\omega}{c} \rho \right)} A_m(\omega) e^{jm\phi_l}, \tag{6.26}$$

being

$$A_m(\omega) = \frac{j}{4} J_m \left(\frac{\omega}{c} \rho \right) H_m^{(1)} \left(\frac{\omega}{c} z \right) e^{jm(\phi_l - \phi_z)}. \tag{6.27}$$

Synthesized acoustic field in frequency domain. Fig. 6.8 shows the real part ($\Re\{\cdot\}$) of the acoustic field reproduced with the proposed technique (6.15) (Fig. 6.8b), with WFS (Fig. 6.8c) and with HOA (Fig. 6.8d). A more quantitative comparison is offered in Fig. 6.9, which shows the approximation error related to the acoustic fields of Fig. 6.8. In particular, the approximation error is defined as

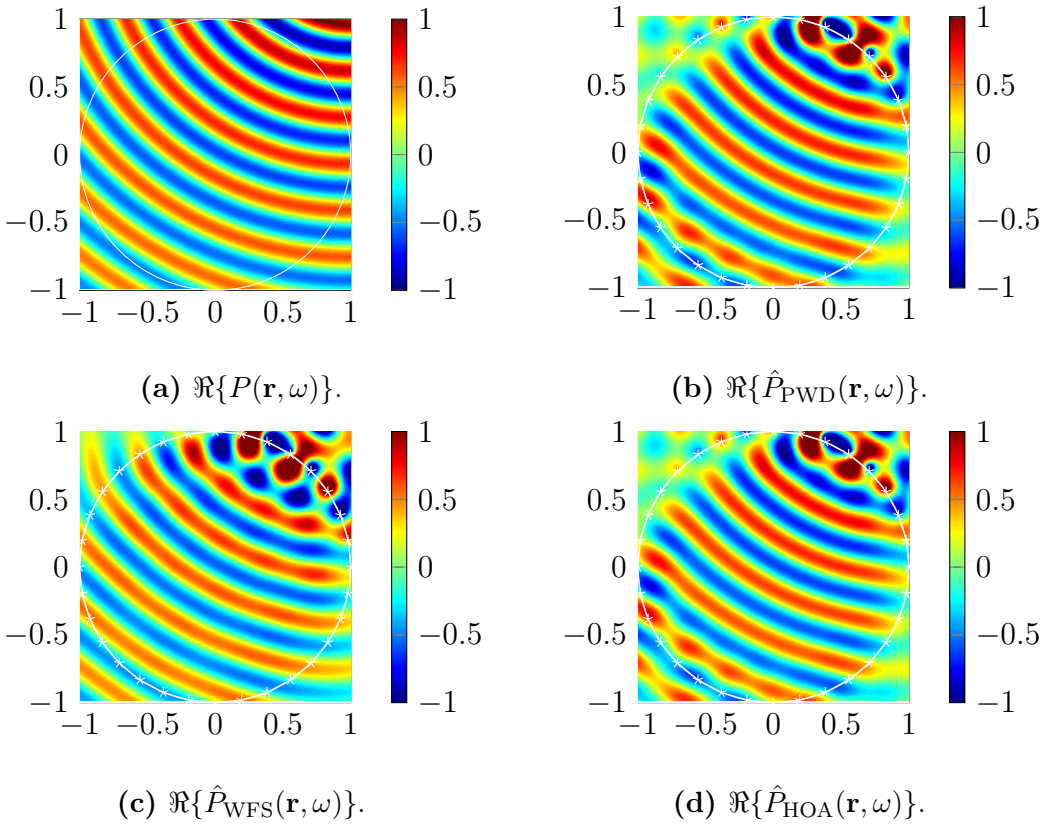
$$\hat{\varepsilon}(\mathbf{r}, \omega) = \frac{|P(\mathbf{r}, \omega) - \hat{P}(\mathbf{r}, \omega)|^2}{|P(\mathbf{r}, \omega)|^2}; \tag{6.28}$$

moreover, $\hat{\varepsilon}_{\text{PWD}}$ denotes the reproduction error for the proposed rendering approach, while $\hat{\varepsilon}_{\text{WFS}}$ and $\hat{\varepsilon}_{\text{HOA}}$ denote the reproduction errors for WFS and HOA, respectively.

Notice that the proposed technique enables an accurate reproduction in a circular area with a radius of approximately 0.7 m and centered in the origin. A very similar spatial distribution of the reproduction error is obtained through the HOA technique presented in Wu and Abhayapala [2009]. On the other hand, WFS exhibits a larger approximation error in the whole area of interest. This is due to the fact that WFS solution is exact only for linear / planar contours $\partial\mathcal{S}$, while the generalization to other contours causes artifacts, as described in Spors and Rabenstein [2006], Spors et al. [2008].

Mean square reproduction error as a function of the virtual source distance. Consider now the rendering performance as a function of the source distance. The target acoustic field is the field due to a 2-D point source in $\mathbf{z} = \rho_z[\cos(\phi_z), \sin(\phi_z)]^T$, $\phi_z = 0$, where ρ_z is made to vary between 1 m and 20 m, i.e. $\rho_z = \{z \in \mathbb{R} | 1 \leq z \leq 20\}$, with the same reproduction setup described above. Fig. 6.10 shows the spatially averaged mean-square reproduction error, i.e.

$$\hat{\varepsilon}(\omega) = \frac{1}{\pi \bar{r}^2} \int_0^{\bar{\rho}} \int_0^{2\pi} \frac{|P(\mathbf{r}, \omega) - \hat{P}(\mathbf{r}, \omega)|^2}{|P(\mathbf{r}, \omega)|^2} \, d\phi_r \, d\rho \tag{6.29}$$



(a) $\Re\{P(\mathbf{r},\omega)\}$. **(b)** $\Re\{\hat{P}_{\text{PWD}}(\mathbf{r},\omega)\}$.

(c) $\Re\{\hat{P}_{\text{WFS}}(\mathbf{r},\omega)\}$. **(d)** $\Re\{\hat{P}_{\text{HOA}}(\mathbf{r},\omega)\}$.

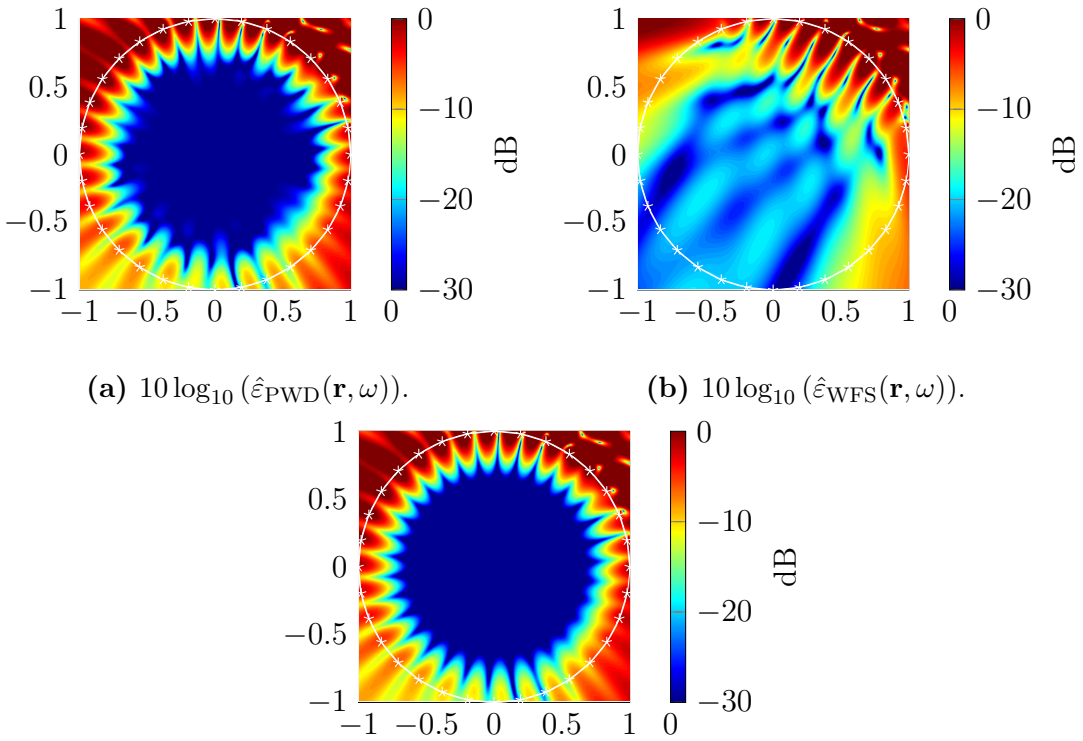
Figure 6.8: Reproduction of the acoustic field with a circular array. Stars mark the location of loudspeakers, while the white circle denotes the contour $\partial\mathcal{S}$. $\omega = 2\pi f$, $f = 1\,\text{kHz}$.

as a function of the source distance ρ_z. The reproduction error has been averaged over a circular area of radius $\bar{\rho} = 0.7\,\text{m}$, which corresponds to the area where the considered rendering techniques exhibit the best performance. An analysis of the results enables the conclusion that for the proposed rendering technique and for HOA the reproduction error is constant with respect to the distance of the source, while for WFS the reproduction error is not constant with source distance. The performance of the proposed method method is similar to the one of HOA; this is due to the fact that, as described in Section 6.3.2, in the proposed method multiple HOA filters are used to implement secondary sources.

6.5.2 Acoustic Field Synthesis with an Elliptical Array

In order to show the applicability of the approach proposed in Section 6.4.1, in this paragraph we show simulations with an elliptical distribution of loudspeakers, as a particular case of a convex curve. In particular, the direction of propagation of plane waves in polar coordinates is $\mathbf{s}(\phi) = (\rho(\phi), \phi)$, being $\rho(\phi) = ab/\sqrt{(b\cos(\phi))^2 + (a\sin(\phi))^2}$, where $a = 0.7\,\text{m}$ and $b = 1\,\text{m}$ are the semi-major and semi-minor axis of the ellipse, respectively. In this example, loudspeakers are placed at positions related to the direction of propagation of plane waves, i.e. the lth loudspeaker is placed at $\mathbf{v}_l = \mathbf{s}(\phi_l)$.

The target acoustic field is that of a point source in $\mathbf{r}_\text{z} = [1, \sqrt{3}]$, i.e. $\rho_\text{z} = 2\,\text{m}$ and $\phi_\text{z} = \pi/3$ (Fig. 6.11a). The acoustic field $P_{\text{PWD}}(\mathbf{r},\omega)$ is synthesized by a discrete



(a) $10\log_{10}\left(\hat{\varepsilon}_{\mathrm{PWD}}(\mathbf{r},\omega)\right)$.

(b) $10\log_{10}\left(\hat{\varepsilon}_{\mathrm{WFS}}(\mathbf{r},\omega)\right)$.

(c) $10\log_{10}\left(\hat{\varepsilon}_{\mathrm{HOA}}(\mathbf{r},\omega)\right)$.

Figure 6.9: Approximation error in the area $\mathcal{S}$ for a circular loudspeaker array rendering the acoustic field of a point source. $\omega = 2\pi f$, $f = 1\,\mathrm{kHz}$.

distribution of plane waves, as described in Section 6.4.1 (Fig. 6.11b). The acoustic field $\hat{P}_{\mathrm{PWD}}(\mathbf{r},\omega)$ is reconstructed with loudspeakers (Fig. 6.11c), where the spatial filter $\mathbf{h}(\phi_n,\gamma)$ is given by (6.23). The parameter λ for Tikhonov regularization in (6.23) is computed as the maximum singular value of the propagation matrix $\mathbf{G}(\mathbf{r}|\mathbf{v}_l,\omega)$ scaled by a factor 0.01; this is a common choice in the design of loudspeaker systems (see, e.g. Olivieri et al. [2013]).

6.5.3 Acoustic Field Synthesis with a Linear Array

Here simulations are shown with a linear deployment of loudspeakers, as discussed in Section 6.4.2. Plane waves propagate with directions $\mathbf{s}(\phi) = [x_0, x_0\tan(\phi)]^T$, where $x_0 = 0.9\,\mathrm{m}$ and ϕ ranges from $\arctan(-y_0/x_0)$ to $\arctan(y_0/x_0)$, $y_0 = 1\,\mathrm{m}$. $L = 32$ loudspeakers are placed to uniformly sample the line segment.

The target acoustic field is that of a point source in $\mathbf{r}_{\mathrm{z}} = [2\,\mathrm{m}, 0]^T$. The acoustic field $P_{\mathrm{PWD}}(\mathbf{r},\omega)$ represents the acoustic field synthesized by the limited set of plane waves considered, as given in (6.20). Fig. 6.12a shows the real part of the acoustic field $P_{\mathrm{PWD}}(\mathbf{r},\omega)$ and Fig. 6.12d shows the reconstruction error with respect to the target field. The acoustic field $\hat{P}_{\mathrm{PWD}}(\mathbf{r},\omega)$ reconstructed with loudspeakers is given by (6.21) (Fig. 6.12b), where the spatial filter $\mathbf{h}(\phi_n,\omega)$ is given by (6.23). The parameter λ, as in the case of the elliptical array, is chosen to be the largest singular value of the propagation matrix scaled by 0.01. The acoustic field reproduced with WFS is obtained from [Spors et al., 2008, Eq. (23)] (Fig. 6.12c). To provide a more quantitative comparison, Fig. 6.12e

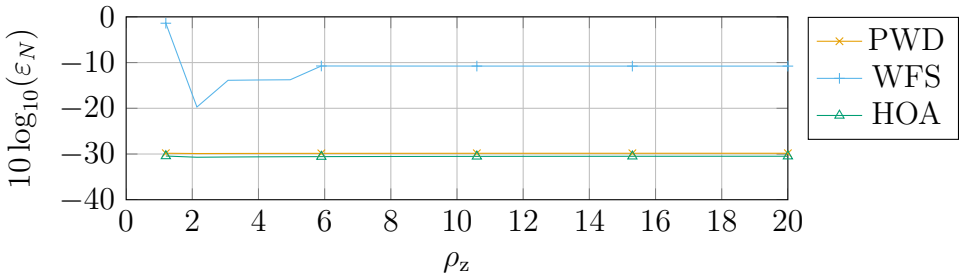


Figure 6.10: Spatially averaged mean-squared error (6.29) as a function of the source distance ρ_z for the considered rendering techniques.

and Fig. 6.12f show the approximation errors. We notice that the proposed technique allows us to obtain a low reproduction error in the area on the left of the loudspeaker array, with similar results with respect to WFS, which provides an optimal solution for the linear geometry.

6.6 Discussion

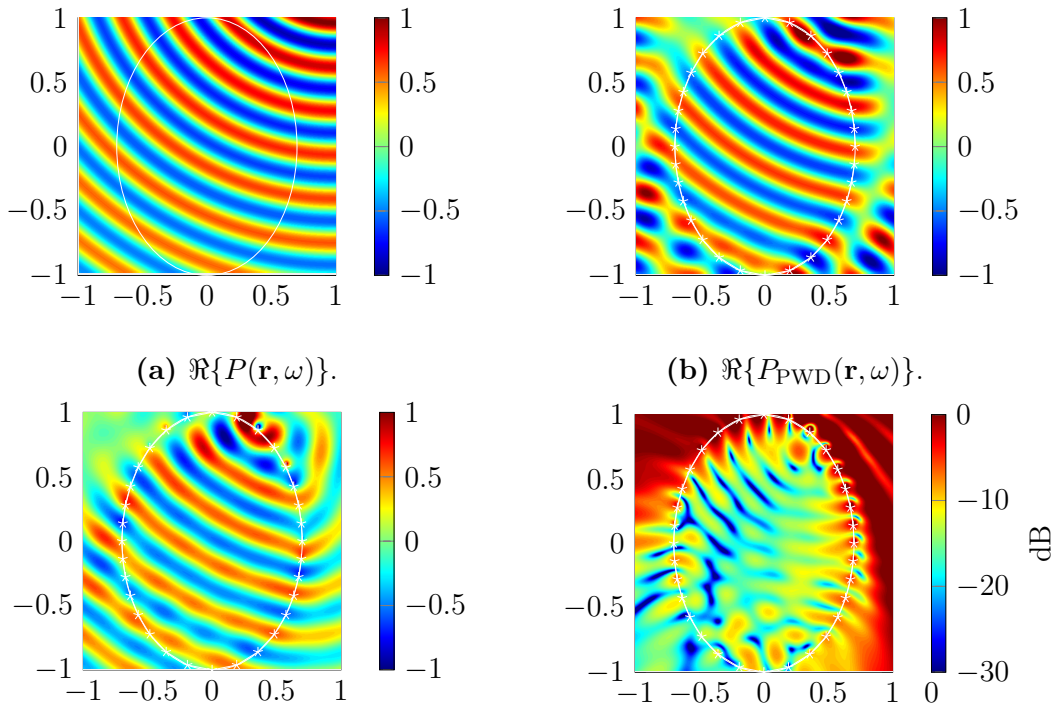
The rendering method presented so far is theoretically appealing and it offers intuitive insights on the directional structure of the acoustic field to be reproduced. Moreover, with respect to other state-of-the-art rendering methods, the loudspeaker driving filters do not have to be re-computed from scratch every time the desired acoustic field varies. Indeed, the possibility of splitting loudspeaker driving filters into two components – one dependent on the actual acoustic field to be reproduced, cfr.(6.4) and (6.8), while the other dependent only on the loudspeaker geometry, cfr. (6.12) and (6.23) – enables implementation with reduced complexity; furthermore, the filter component independent on the actual desired acoustic field can be optimized with respect to real-world operating conditions.

Consider the simplest scenario of a circular loudspeaker deployment. In many practical situation, the assumptions introduced in order to derive the proposed rendering method are not realistic. Indeed, most common loudspeakers available today are not suitable to generate the acoustic field in (3.61), as it is required by the proposed method. In this case, mismatches arise when one tries to use real world loudspeakers (whose propagation function is more close to the model of a point source (3.59), rather than an height invariant source (3.61)) with the proposed rendering method. It should be remarked that this consideration is valid for all state-of-the-art acoustic rendering systems, not just for the proposed one; e.g., in the context of WFS, adaptations have been proposed in order to compensate for the propagation mismatch between (3.59) and (3.61) imposing that the acoustic field is correct at a reference point inside the listening area [Spors et al., 2008, Sec. 4.2], while accepting artifacts at other listening positions.

Assuming that a compensation technique of this sort is available, there is still another issue that remains unsolved. This issue is related to the power that the loudspeaker array must emit in order to correctly synthesize plane wave components; this quantity is proportional to the energy of the driving filters (6.12) for each loudspeaker, given by

$$Q(\omega) = 10\log_{10}\left(\sum_{l=1}^{L}|h_l(\phi_n,\omega)|^2\right) = 10\log_{10}\left(\frac{4}{L}\sum_{l=1}^{L}\left|\sum_{m=-M}^{M}\frac{e^{jm(\phi_l-\phi_n)}}{H_m^{(1)}\left(\frac{\omega}{c}\rho\right)}\right|^2\right). \quad (6.30)$$

It is easy to verify that $Q(\omega)$ is independent on the DoA ϕ_n of the plane wave to be



(a) $\Re\{P(\mathbf{r},\omega)\}$.

(b) $\Re\{P_{\mathrm{PWD}}(\mathbf{r},\omega)\}$.

(c) $\Re\{\hat{P}_{\mathrm{PWD}}(\mathbf{r},\omega)\}$.

(d) $10\log_{10}\left(\hat{\varepsilon}_{\mathrm{PWD}}(\mathbf{r},\omega)\right)$.

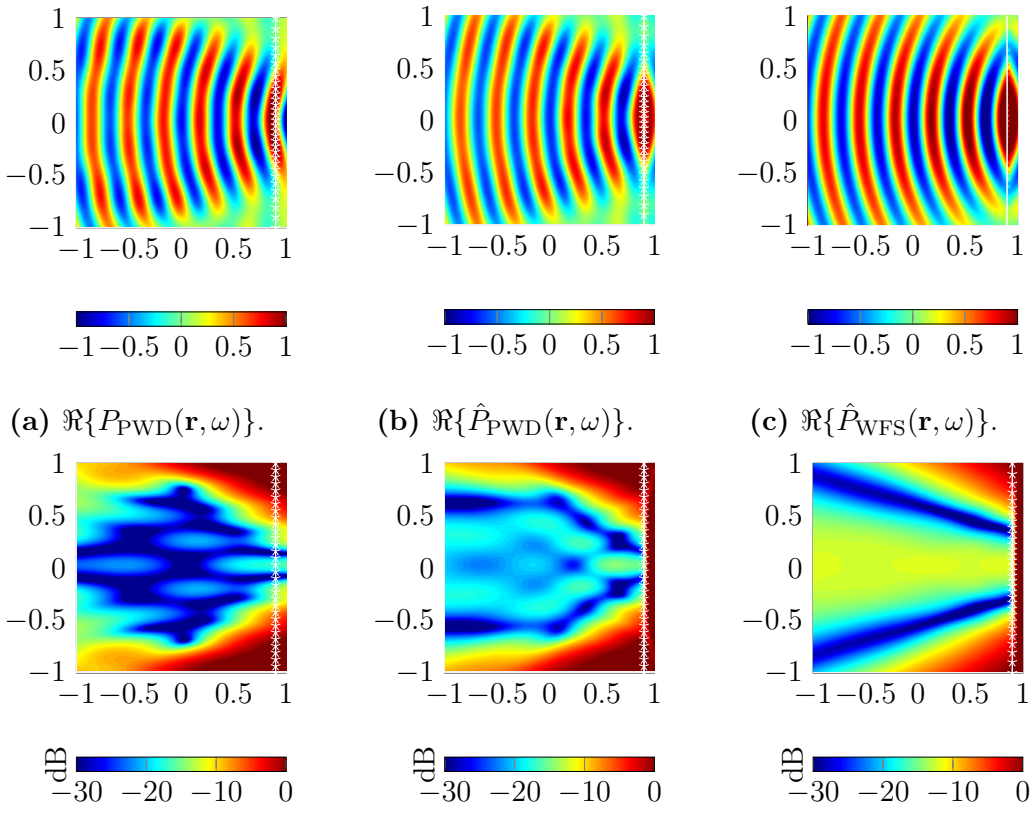
Figure 6.11: Acoustic field reproduction with an elliptical distribution of secondary sources. Stars mark the location of secondary sources, while the white ellipse denotes the contour $\partial\mathcal{S}$. $\omega = 2\pi f$, $f = 1\,\mathrm{kHz}$.

synthesized. Figure 6.13 shows the quantity $Q(\omega)$ as a function of the temporal frequency and radius ρ of the reproduction area. One can easily observe that the required power is almost zero at low temporal frequencies: this leads to robustness issues, since even small excitation signals at low frequencies make the system unstable.

Another assumption that is not completely justified in real-world situations is that, in general, loudspeakers employed in an array are not identical, yielding to propagation functions that are different on a loudspeaker basis. This is mainly due to the tolerances in the loudspeaker manufacturing process, which lead to significant mismatches between the ideal loudspeaker propagation and the actual one.

For these reasons, the theoretical solution for the rendering method derived in the previous sections may not provide optimal results when employed in real-world scenarios. Indeed, this solution, as well as state-of-the-art analytic rendering solutions, may not be robust to mismatches in the propagation conditions.

A similar situation arises when considering the least-squares solution (6.23) for the rendering of plane waves with non-circular loudspeaker deployments. In this case, (6.22) uses a diagonal loading (Tikhonov regularization) of the propagation matrix in order to regularize the solution and improve its robustness to modeling mismatches. However, the selection of an optimal regularization parameter λ is not a trivial task and a general solution does not exist. Even the criterion employed for the simulations presented in the previous sections appears questionable in many scenarios, since it has to be scaled by a factor whose selection is crucial for the successful application of the solution. From an engineering perspective, a regularization approach that does not depend heavily on the



(a) $\Re\{P_{\text{PWD}}(\mathbf{r},\omega)\}$. **(b)** $\Re\{\hat{P}_{\text{PWD}}(\mathbf{r},\omega)\}$. **(c)** $\Re\{\hat{P}_{\text{WFS}}(\mathbf{r},\omega)\}$.

(d) $10\log_{10}\left(\varepsilon_{\text{PWD}}(\mathbf{r},\omega)\right)$. **(e)** $10\log_{10}\left(\hat{\varepsilon}_{\text{PWD}}(\mathbf{r},\omega)\right)$. **(f)** $10\log_{10}\left(\hat{\varepsilon}_{\text{WFS}}(\mathbf{r},\omega)\right)$.

Figure 6.12: Acoustic field reproduction with a linear distribution of secondary sources. Stars mark the location of secondary sources, while the white line denotes the contour $\partial\mathcal{S}$. $\omega = 2\pi f$, $f = 1\,\text{kHz}$.

choice of a parameter would be of great interest.

Next sections will show how that the most impacting non-idealities can be modeled as an additive random variable. A parameter-free regularization technique will be presented in the following section, exploiting the statistical knowledge of the loudspeaker propagation function in order to provide a solution that is robust to modeling mismatches. This solution is obtained by casting the filter design problem (6.22) in the form of a Robust Least Squares (RLS) problem, whose solution is known from the literature on convex optimization; precisely, the solution is described in Boyd and Vandenberghe [2009, p. 318, Sec. 6.4].

6.7 Robust Parameter-Free Regularization

Consider the least squares problem (6.23) for the computation of loudspeaker driving filters able to reproduce plane waves with non-circular loudspeaker deployments. It has been noted in Mabande et al. [2009], that a direct solution to (6.23) suffers from relevant errors when the knowledge of the propagation functions is not exact. In particular, the propagation functions from loudspeakers to control points is affected by uncertainties due to manufacturing tolerances, loudspeaker positioning errors, modeling inaccuracies, etc.

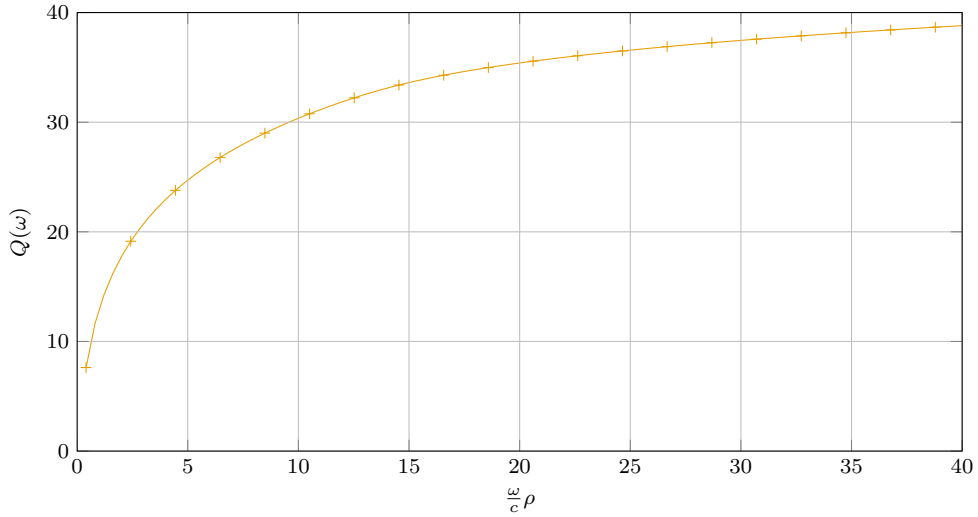


Figure 6.13: Energy $Q(\omega)$ of the filter coefficients (6.12) as a function of $(\omega/c)\rho$. The truncation mode is set to $M = 15$ end the number of loudspeakers is $L = 32$.

Several approaches have been proposed to address this sensitivity problem. In Mabande et al. [2009] the authors present a method that incorporates a constraint for the White Noise Gain (WNG) into a least-squares beamformer design and still leads to a convex optimization problem that can be solved directly. An extension to time-domain design has been presented in Mabande et al. [2011]. In Lai et al. [2011] it is stated, however, that it is difficult to select an appropriate level of WNG for any given set of errors in loudspeaker speaker positions and / or directivity pattern; here the authors adopt a Farrow structure for the beamforming and they incorporate the probability density functions for the transducer position error into the design formulation. A beamforming technique that is inherently robust against errors in the propagation matrix is presented in Trucco et al. [2006] for microphone arrays. In Rübsamen and Gershman [2010, 2012], Rübsamen and Pesavento [2012], in the context of mitigating the effect of interferes, the authors extend the 1D covariance matrix fitting approach [Yardibi et al., 2010] to multiple dimensions, and the steering vectors are modeled by means of uncertainty sets. In Doclo and Moonen [2003] two design procedures are proposed, aimed at increasing the robustness of a microphone array design: the first embeds the probability density functions of the steering vectors, whereas the second optimizes the worst-case performance through the minimax criterion. In Levin et al. [2013] the authors propose a technique that modifies the classical loading scheme incorporating a non-diagonal elements to attenuate the effect of sensor uncertainties. The method in Levin et al. [2013], however, is data-dependent, i.e. the filters depend on the signal to be rendered, thus making its use for rendering purposes quite cumbersome.

In this section, the uncertainty in the propagation matrix is modeled as the sum of a deterministic (nominal) component and a random one. This model enables the exploitation of the statistical properties of the random component in order to regularize the solution to the least square problem (6.23), as in the RLS problem discussed in Boyd and Vandenberghe [2009, p. 318, Sec. 6.4]. A non-diagonal loading of the original least squares problem is obtained as solution.

6.7.1 Problem Statement

The solution (6.22) does not take into account uncertainty or variations in speaker locations and their directivity pattern, which constitute a main issue in loudspeaker array design [Tashev et al., 2008]. Notice that the propagation function in (6.22) contains contributions related to both the attenuation with distance from the loudspeakers and the loudspeaker directivity pattern. Therefore, the propagation function can be factorized as

$$G(\mathbf{r}|\mathbf{v}_l,\omega) = G_0(\mathbf{r}|\mathbf{v}_l,\omega)d(\mathbf{r}|\mathbf{v}_l,\omega), \tag{6.31}$$

where the term $d(\mathbf{r}|\mathbf{v}_l,\omega)$ only contains the directivity pattern, hence it does not depend on the absolute relative position of $\mathbf{r}$ and $\mathbf{v}_l$ but just on their relative angle

$$\gamma = \arccos\left(\frac{\langle \mathbf{r},\mathbf{v}_l\rangle}{\|\mathbf{r}\|\cdot\|\mathbf{v}_l\|}\right), \tag{6.32}$$

and will be denoted in the following by $d_l(\omega,\gamma)$. The symbol $G_0(\cdot)$ denotes the free-field Green's function (3.59).

Here, the uncertainty in loudspeaker directivity pattern is modeled as a random variable $\delta(\omega,\gamma)$ of variance σ_D^2, which alters the nominal directivity pattern $\overline{d}(\omega,\gamma)$ usually specified by the loudspeaker manufacturer. Thus, it can be written

$$d_l(\omega,\gamma) = \overline{d}(\omega,\gamma) + \delta(\omega,\gamma), \tag{6.33}$$

where $\delta(\omega,\gamma) \sim \mathcal{N}(0,\sigma_D^2(\omega,\gamma))$. Notice that the same approach could also be used for the modeling of the uncertainty of the speaker locations. This aspect, however, goes beyond the scope of this section. By substituting (6.33) into (6.31), one obtains an expression that relates the uncertainty on propagation function with the uncertainty on the directivity pattern

$$G_l(\omega,\gamma) = G_0(\mathbf{r}|\mathbf{v}_l,\omega)\overline{d}(\omega,\gamma) + G_0(\mathbf{r}|\mathbf{v}_l,\omega)\delta(\omega,\gamma), \tag{6.34}$$

or, in matrix notation,

$$\mathbf{G}(\omega) = \mathbf{G}_0(\omega)\odot\overline{\mathbf{D}}(\omega) + \mathbf{G}_0(\omega)\odot\mathbf{\Delta}(\omega) \tag{6.35}$$

where $[\mathbf{G}_0(\omega)]_{q,l} = G_0(\mathbf{r}|\mathbf{v}_l,\omega)$ for γ_q given in (6.32), $[\overline{\mathbf{D}}(\omega)]_{q,n} = \overline{d}(\omega,\gamma_q)$, $[\mathbf{\Delta}(\omega)]_{q,n} = \delta(\omega,\gamma_q)$ and $\odot$ denotes Hadamard product. Finally, $\mathbf{G}(\omega)$ is expressed as the random variable

$$\mathbf{G}(\omega) = \overline{\mathbf{G}}(\omega) + \mathbf{U}(\omega), \tag{6.36}$$

where $\overline{\mathbf{G}}(\omega) = \mathbf{G}_{FF}(\omega)\odot\overline{\mathbf{D}}(\omega)$ is the mean value of $\mathbf{G}(\omega)$ and $\mathbf{U}(\omega) = \mathbf{G}_{FF}(\omega)\odot\mathbf{\Delta}(\omega)$ describes its statistical variation.

Next paragraphs will describe how a knowledge of the statistical model of $\mathbf{G}(\omega)$ can be exploited to derive a closed-form solution for the problem of designing a robust loudspeaker driving filter.

6.7.2 Robust Least Squares Solution

Consider the least squares problem in (6.22) and modify it for the purpose of minimizing the average approximation error, i.e.

$$\underset{\hat{\mathbf{h}}(\phi_n,\omega)}{\arg\min}\, E[\|\mathbf{G}(\omega)\hat{\mathbf{h}}(\phi_n,\omega) - \mathbf{d}(\phi_n,\omega)\|_2^2], \tag{6.37}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. Recalling the factorization of the propagation matrix in (6.36) one can express the objective function as [Boyd and Vandenberghe, 2009, p. 318]

$$\begin{aligned} E[\|\mathbf{G}\hat{\mathbf{h}} - \mathbf{d}\|_2^2] &= E[(\overline{\mathbf{G}}\hat{\mathbf{h}} - \mathbf{d} + \mathbf{U}\hat{\mathbf{h}})^H(\overline{\mathbf{G}}\hat{\mathbf{h}} - \mathbf{d} + \mathbf{U}\hat{\mathbf{h}})] \\ &= (\overline{\mathbf{G}}\mathbf{h} - \mathbf{d})^H(\overline{\mathbf{G}}\mathbf{h} - \mathbf{d}) + E[\hat{\mathbf{h}}^H\mathbf{U}^H\mathbf{U}\hat{\mathbf{h}}] \\ &= \|\overline{\mathbf{G}}\hat{\mathbf{h}} - \mathbf{d}\|_2^2 + \hat{\mathbf{h}}^H\mathbf{P}\hat{\mathbf{h}}, \end{aligned} \tag{6.38}$$

where the dependency on ω and ϕ_n has been dropped for compactness of the notation, and $\mathbf{P} = E[\mathbf{U}^H\mathbf{U}]$. Hence, the statistical robust least-squares problem can be written as a regularized least-squares problem

$$\arg\min_{\hat{\mathbf{h}}} \|\overline{\mathbf{G}}\hat{\mathbf{h}} - \mathbf{d}\|_2^2 + \|\mathbf{P}^{1/2}\hat{\mathbf{h}}\|_2^2, \tag{6.39}$$

whose solution is [Boyd and Vandenberghe, 2009, p. 319]

$$\hat{\mathbf{h}} = (\overline{\mathbf{G}}^H\overline{\mathbf{G}} + \mathbf{P})^{-1}\overline{\mathbf{G}}^H\mathbf{d}. \tag{6.40}$$

It is important to observe that (6.40) corresponds to a loading of the least-squares solution through matrix $\mathbf{P}$. With respect to the Tikhonov regularization, however, the loading is not diagonal, as $\mathbf{P}$ may have non-diagonal elements different from zero. A similar result was attained in Levin et al. [2013] in the context of data-dependent beamformers for microphone arrays.

6.7.3 Validation of the Robust Regularization Technique

The simulations and experimental results reported in this paragraph are intended to provide a validation for the robust regularization technique presented in the previous paragraph. The validation is conducted in a significant illustrative setting, where the reproduction of a broadband plane wave is in order. Hence, it is desirable that the synthesized acoustic field exhibits invariant characteristics over the temporal frequency range of interest.

Spatial filters able to synthesize broadband and frequency-invariant plane waves have been first introduced in Ward et al. [1995], Abhayapala et al. [2000] based on the spatial Fourier transform of a continuous aperture. Analytic approaches have been presented for specific discrete array deployments, e.g. linear [Sekiguchi and Karasawa, 2000], cylindrical [Teutsch and Kellermann, 2005], spherical [Meyer and Elko, 2002, Abhayapala and Ward, 2002, Rafaely, 2005]. To overcome the restrictions on transducer locations imposed by analytical approaches, in Parra [2006] a numerical method based on the solution of a least-squares problem is introduced, originally conceived for microphone arrays but easily applicable to loudspeaker arrays. This is a two-step procedure, which first solves numerically a least-squares problem that decouples the frequency-behavior from the spatial-behavior through a change of basis, and then solves a second least-squares problem for the design of the spatial filter.

Here an extension of the data-independent beamformer in Parra [2006] is adopted, using the proposed robust regularization technique in both problems of change of basis and spatial filtering.

In order to produce a frequency invariant plane wave, in Parra [2006] a basis transformation is proposed, which leads from $\mathbf{G}(\omega)$ to a new basis $\widetilde{\mathbf{G}}$ through a transformation matrix $\mathbf{B}(\omega)$, which minimizes the approximation error in the least squares sense

$$\arg\min_{\mathbf{B}(\omega)} \|\mathbf{G}(\omega)\mathbf{B}(\omega) - \widetilde{\mathbf{G}}\|_2^2]. \tag{6.41}$$

In Parra [2006] the new basis $\widetilde{\mathbf{G}}$ is chosen to be frequency-invariant. For this reason, the new basis consists of spherical harmonics (cfr. 3.3.2) and the approximation is truncated to order L and degree M (cfr. 4.2.3). The propagation matrix in the direction γ in the new basis is thus written as

$$[\widetilde{\mathbf{G}}] = [Y_0^0(\gamma), Y_1^{-1}(\gamma), \ldots, Y_L^M(\gamma)]. \tag{6.42}$$

The design of the spatial filter, in Parra [2006], is accomplished in the transformed domain. In particular, the spatial filter is computed to approximate an objective array response $\mathbf{d}$ in the least squares sense as

$$\arg\min_{\widetilde{\mathbf{h}}(\omega)} \|\mathbf{G}(\omega)\mathbf{B}(\omega)\widetilde{\mathbf{h}}(\omega) - \mathbf{d}\|_2^2] \tag{6.43}$$

and the filter coefficients in the original domain are obtained as

$$\mathbf{h}(\omega) = \mathbf{B}(\omega)\widetilde{\mathbf{h}}(\omega). \tag{6.44}$$

However, the derivation in Parra [2006] and briefly summarized here does not take into account possible uncertainties in the loudspeaker propagation function. In order to take into account this uncertainty, we follow the procedure presented in Sec. 6.7.2. In Parra [2006] the propagation matrix appears in both steps of basis transformation and spatial filter design.

Consider first the basis transformation problem in (6.41), with the objective of minimizing the mean case approximation error. Exploiting the factorization of the propagation matrix in (6.31), the problem can be written as a regularized least-squares problem, whose solution is [Boyd and Vandenberghe, 2009]

$$\mathbf{B} = \left(\overline{\mathbf{G}^H\mathbf{G}} + \mathbf{P}\right)^{-1}\overline{\mathbf{G}}^H\widetilde{\mathbf{G}}. \tag{6.45}$$

Consider now the problem in (6.43) for the computation of the beamforming filters. Upon applying the same procedure, one obtains

$$\widetilde{\mathbf{h}} = (\overline{\mathbf{GB}}^T\overline{\mathbf{GB}} + \mathbf{\Pi})^{-1}\overline{\mathbf{GB}}^T\mathbf{f}, \tag{6.46}$$

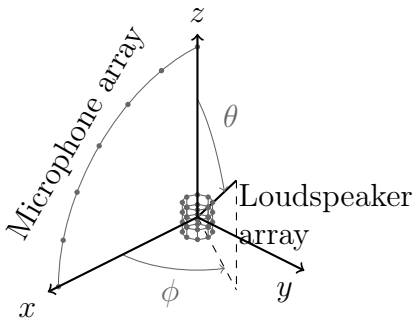
where $\overline{\mathbf{GB}} = \overline{\mathbf{G}}\mathbf{B}$, $\mathbf{\Upsilon} = \mathbf{UB}$ and

$$\mathbf{\Pi} = E[\mathbf{\Upsilon}^T\mathbf{\Upsilon}] = \mathbf{B}^T\mathbf{P}\mathbf{B}. \tag{6.47}$$

Finally, the filter coefficients $\mathbf{h}$ in the original domain are obtained by the change of basis in (6.44).

It is important to observe that both (6.46) and (6.45) correspond to a loading of the least squares solution through matrices $\mathbf{P}$ and $\mathbf{\Pi}$, respectively. With respect to the Tikhonov regularization, however, the loading is not diagonal, as $\mathbf{P}$ and $\mathbf{\Pi}$ may have non-diagonal elements different from zero. Notice also, from (6.47), that $\mathbf{P}$ and $\mathbf{\Pi}$ are related through $\mathbf{B}$. Therefore, with respect to the solution presented in Parra [2006], the only additional knowledge required for a complete characterization of the proposed robust spatial filters is the matrix $\mathbf{P}$, determined by the statistics of the loudspeaker directivity patterns.

(a) Loudspeake49 array.



(b) Measurement setup.

Figure 6.14: Custom cylindrical loudspeaker array (Fig. 6.14a) and measurement setup (Fig. 6.14b).

6.7.4 Setup for the Experimental Validation.

The experimental validation is conducted in a room with short reverberation time, being $T_{60} \approx 50\,\text{ms}$ in the frequency band $200\,\text{Hz} \div 5\,\text{kHz}$. For illustrative purposes, we adopt a custom cylindrical loudspeaker array, designed by B&C Speakers, mounting $N = 32$ 2-inch full-range drivers; the spatial filtering problem is here considered in a full 3-D setting. The radius of the cylinder is 9.2 cm, while the distance between drivers along the z axis is 7.5 cm. This results in a maximum radiation mode $L = 3$, having Nyquist frequency $f_{\text{Nyq},3} \approx 2.2\,\text{kHz}$ (cfr. Sec. 4.2.3). Fig. 6.14a shows the loudspeaker array.

The measurement setup, along with the reference frame, is shown in Fig. 6.14b. Notice that the origin of the reference frame coincides with the center of the loudspeaker array. In order to estimate the response of the array of loudspeakers, $M = 7$ measurement microphones are deployed at angles $\gamma = (\theta_t, 0), t = 1, \ldots, M$ and distance fixed and equal to 1.3 m, thus forming an arc of sensors on the xz plane. The loudspeaker array is mounted on a stepper turntable, which enable to rotate the array in the xy plane towards directions $\phi_p,\ p = 1, \ldots, Q_\phi$. The microphones have been calibrated in such a way that their response to an omnidirectional source located in the center of the reference frame is equal in amplitude for all the $M = 7$ microphones. For each rotation angle of the turntable and for each microphone, the response of the array in the direction (θ_q, ϕ_p) is evaluated. As a consequence, the response is evaluated on a grid of $\gamma_q = (\phi_p, \theta_t)$ propagation directions, indexed by $q = Q_\phi(t-1) + p$, with resolution $180°/(Q_\theta - 1) = 15°$ for the elevation angle and $360°/Q_\phi = 5°$ for the azimuth angle.

In all simulations and experiments shown in this section, the frequency-independent desired response of the loudspeaker array $\mathbf{d}(\omega)$ is a plane wave in direction $\gamma = (90°, 0°)$

$$\begin{cases} f(\omega, \gamma_q) = f(\gamma_q) = 1, & \text{if } q = Q_\phi \left(\frac{Q_\theta + 1}{2} - 1 \right) + 1, \\ f(\omega, \gamma_q) = f(\gamma_q) = 0, & \text{otherwise.} \end{cases} \tag{6.48}$$

We choose to truncate the spherical harmonics expansion to $L = 12$ in order to ensure a more numerically accurate basis transformation; we remark that this choice is not intended to increase the accuracy of the reproduced acoustic field since radiation modes higher than $L = 3$ are not reproducible by our setup.

The time-domain filters $h_n(t), n = 1, \ldots, N$ are derived from (6.44) through length-K Inverse Fourier Transform, being $K = 512$. The excitation signals are Golay complementary sequences $a(t)$ and $b(t)$ of length $L_G = 4096$ [Foster, 1986]. Let $r_a(t, \gamma_q) = a(t) * f(t, \gamma_q)$

and $r_b(t,\gamma_q) = b(t) * f(t,\gamma_q)$ be the array responses in direction γ_q due to inputs $a(t)$ and $b(t)$, respectively. Through the definition of Golay complementary sequences [Foster, 1986], the array response is obtained by

$$f(t,\gamma_q) = (1/(2L_G))\,(a(t)r_a(t,\gamma_q) + b(t)r_b(t,\gamma_q))\,, \tag{6.49}$$

which is transformed through length-K Fourier transform to obtain the array response $f(\omega,\gamma_q)$ in the frequency domain. For each rotation p of the turntable the set $d(\omega,\gamma_q), q = Q_\phi(t-1)+p, t = 1,\ldots,M$ of array responses is acquired, such that after a complete rotation of the turntable the array response is sampled on the whole northern hemisphere. Due to the symmetry of the measurement setup and of the loudspeaker array, the response on the southern hemisphere is obtained by symmetrizing the response in the northern hemisphere, i.e.

$$d(\omega,\gamma_q)|_{\substack{q=Q_\phi(t-1)+p,\\ t=M+1,\ldots,Q_\theta}} = d(\omega,\gamma_q)|_{\substack{q=Q_\phi(t-1)+p,\cdot\\ t=M-1,\ldots,1}} \tag{6.50}$$

In the following paragraph, the behavior of the proposed spatial filter is investigated, considering as a figure of merit, the far field array response $\hat{\mathbf{f}}(\omega)$, which for the desired plane wave to be synthesized ideally should be a unit impulse in the direction γ. We compare the array response $\hat{\mathbf{f}}_{\text{tik}}(\omega)$ produced by filters computed from (6.43) adopting Tikhonov regularization, as suggested in Parra [2006], with the array response $\hat{\mathbf{f}}_{\text{rls}}(\omega)$ produced by filters computed from (6.46). The coefficient for the Tikhonov regularization has been set to $-20\,\text{dB}$. For the proposed robust design, we set the variance of the loudspeaker directivity pattern to $\sigma_D^2(\omega,\gamma) = -20\,\text{dB}$. The nominal directivity pattern $\overline{d}(\omega,\gamma)$, and the variance $\sigma_D^2(\omega,\gamma)$ used in both design methodologies, have been experimentally determined by measuring in a preliminary stage individual drivers mounted in the array structure.

6.7.5 Experimental Results

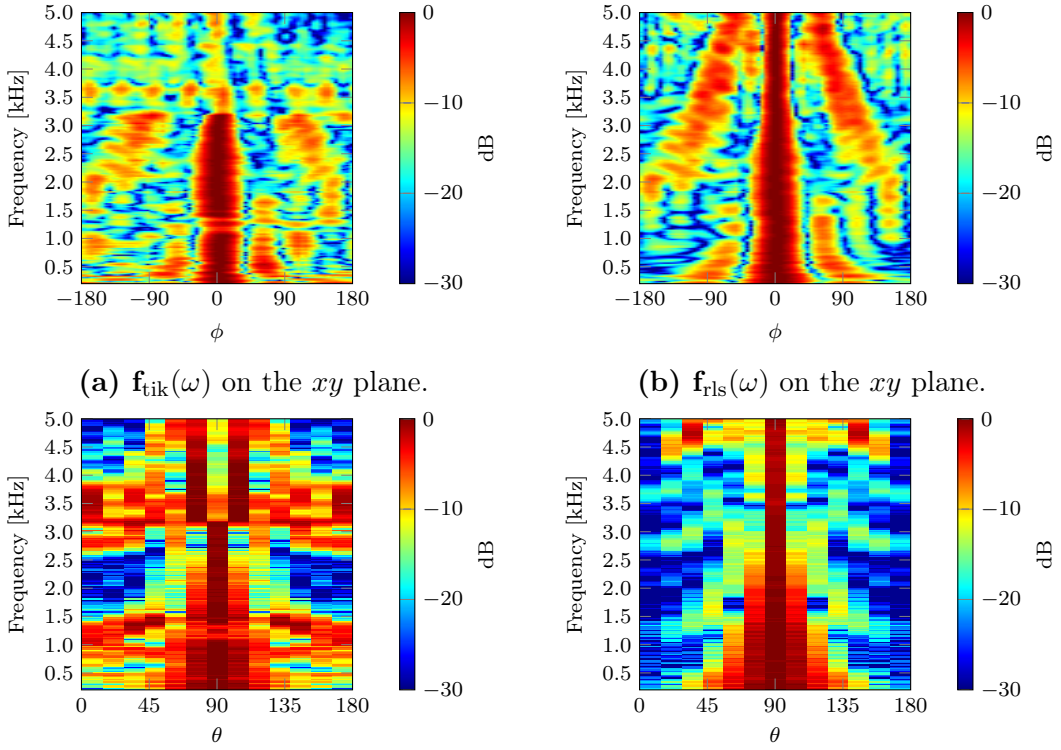
Figs. 6.15a and 6.15b show the array responses on the xy plane, i.e. $\hat{f}_{\text{tik}}(\omega,(90°,\phi_p))$ and $\hat{f}_{\text{rls}}(\omega,(90°,\phi_p))$, $p = 1,\ldots,Q_\phi$.

Notice that with the proposed design methodology the beam is reasonably rendered in the whole frequency band of interest, despite the presence of aliasing starting from the Nyquist frequency. On the other hand, with Tikhonov regularization, the beam is not rendered for frequencies above 3 kHz.

Figs. 6.15c and 6.15d show the array responses on the xz plane, i.e. $\hat{f}_{\text{tik}}(\omega,(\theta_t,0°))$ and $\hat{f}_{\text{rls}}(\omega,(\theta_t,0°))$, $t = 1,\ldots,Q_\theta$. As already observed for the xy plane, also in this case $\hat{\mathbf{f}}_{\text{rls}}(\omega)$ clearly exhibits a reasonably narrow beam, though widened at low frequencies due to the small number of loudspeakers available to control the sound beam along the direction of the elevation. On the other hand, $\hat{\mathbf{f}}_{\text{tik}}(\omega)$ only matches the desired response in a small frequency range between 1.5 kHz and 3 kHz.

In order to quantitatively assess the performance of the proposed design methodology, Fig. 6.16 shows the DI of the loudspeaker array as a function of frequency. The DI is denoted by $\text{DI}(\omega)$ and it is defined as the ratio between the power radiated in the solid angle towards which the beam is steered and the average power radiated on the sphere, as in Trees [2002, p. 60, Eq. 2.144]

$$\text{DI}(\omega) = 10\log_{10}\left[\frac{\hat{f}(\omega,(90°,0°))}{1/4\pi\int_0^\pi\int_0^{2\pi}\sin(\theta)\hat{f}(\omega,(\theta,\phi))\,d\phi\,d\theta}\right]. \tag{6.51}$$



(a) $\mathbf{f}_{\text{tik}}(\omega)$ on the xy plane.

(b) $\mathbf{f}_{\text{rls}}(\omega)$ on the xy plane.

(c) $\mathbf{f}_{\text{tik}}(\omega)$ on the xz plane.

(d) $\mathbf{f}_{\text{rls}}(\omega)$ on the xz plane.

Figure 6.15: Measured array response on the xy plane (Figs. 6.15a and 6.15b) and on the xz plane (Figs. 6.15c and 6.15d). The amplitude response is normalized on a frequency basis and represented in a dB scale.

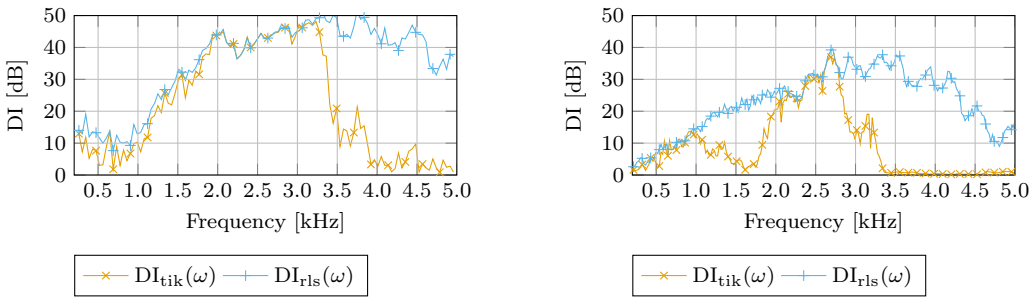
Fig. 6.16a shows the DI computed in a simulative setup that reproduces the real one, while Fig. 6.16b plots the DI for real data. Both results clearly show that the energy of $\hat{\mathbf{f}}_{\text{tik}}(\omega)$ is similar to an omnidirectional pattern at frequencies above 3.5 kHz, while a significant portion of the energy of $\hat{\mathbf{f}}_{\text{rls}}(\omega)$ still remains concentrated in the desired region.

The robustness of the proposed design methodology is highlighted through the analysis of the power of filter coefficients, as defined in (6.30). Fig. 6.17 compares the power of the filters computed with Tikhonov regularization with the power of the filters computed from (6.46). Notice that the proposed methodology is able to guarantee an improved robustness in the whole frequency band of interest, avoiding the stability problem of Tikhonov regularization outside the mid frequency band.

6.8 Regularized Acoustic Field Synthesis

This sections shows the use of the regularization technique introduced in Sec. 6.7.2 for the purpose of making the filter design operation in (6.22) more robust to uncertainties in the loudspeaker propagation function. As it has been recalled in past paragraphs, those uncertainties may arise from a mismatch between the ideal directional behavior of the loudspeakers (i.e. variations from the omnidirectional behavior assumed in the derivation of the rendering method proposed in this chapter), or even from a non-ideal positioning of the loudspeakers, whose position may differ from the ideal one due to positioning errors.

In the following, simulations are shown to assess the use of the proposed regularization



(a) Simulated. **(b)** Experimental.

Figure 6.16: Simulated (Fig. 6.16a) and measured (Fig. 6.16b) DI, obtained with Tikhonov regularization ($\mathrm{DI}_{\mathrm{tik}}(\omega)$, blue curves) and with the proposed design methodology ($\mathrm{DI}_{\mathrm{rls}}(\omega)$, red curves).

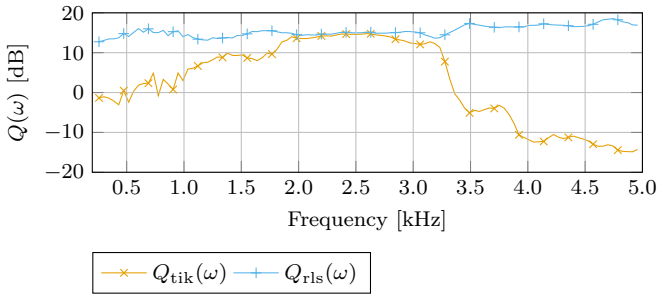


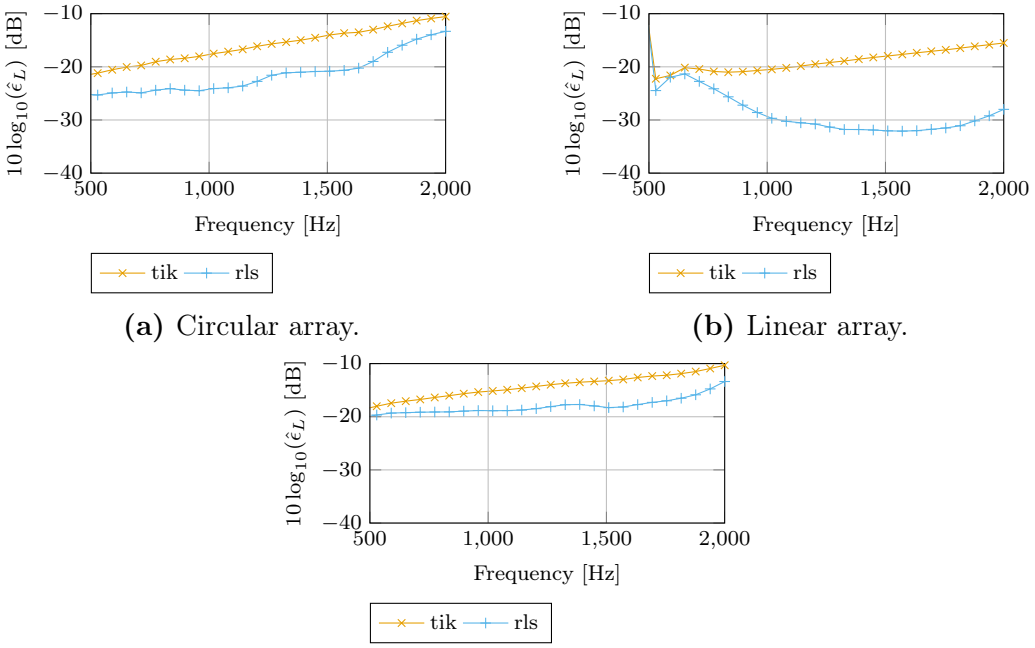
Figure 6.17: White Noise Gain resulting from filters computed with Tikhonov regularization ($Q_{\mathrm{tik}}(\omega)$, blue curve) and from filters computed with the proposed methodology ($Q_{\mathrm{rls}}(\omega)$, red curve).

technique in the context of the acoustic field synthesis method introduced in this chapter. In all the cases that are shown, the nominal propagation function is perturbed with an additive random term, as in (6.35), drawn from a Gaussian distribution with zero mean and variance $\sigma_D^2 = 0.001$. The robust regularization term $\mathbf{P}$ has been computed as the mean of 100 realizations of the random process $\boldsymbol{\Delta} \sim \mathcal{N}(0, \sigma_D^2)$. The Tikhonov regularization parameter λ is chosen to be the maximum singular value of the propagation matrix $\mathbf{G}$, scaled by 0.01 as suggested in Olivieri et al. [2013]. All the plots shown in this section feature a blue curve showing the error obtained with Tikhonov regularization for filter design, as in (6.23); on the other hand the red curve shows the error obtained with the proposed regularization technique (6.40).

In particular, Fig. 6.18a shows the spatially averaged mean-squared error (6.29) as a function of temporal frequency for a circular array composed by $L = 52$ equiangularly spaced loudspeakers. The desired field is that of a 2-D point source at $\rho_{\mathrm{z}} = 2$ and $\phi_{\mathrm{z}} = \pi/3$; the acoustic field is band-limited to $M = 25$.

Figure 6.18b shows the spatially averaged mean-squared error (6.29) as a function of temporal frequency for a linear array composed by $L = 32$ loudspeakers lying on the axis $x = 1\,\mathrm{m}$ between $y_0 = 1\,\mathrm{m}$ and $-y_0 = -1\,\mathrm{m}$. The desired field is that of a 2-D point source at $\rho_{\mathrm{z}} = 2\,\mathrm{m}$, $\phi_{\mathrm{z}} = \deg 0$; even in this case the acoustic field is band-limited to $M = 25$.

Finally, Fig. 6.18c shows the spatially averaged mean-squared error (6.29) as a function of temporal frequency for an elliptical loudspeaker array composed by $L = 52$ loudspeakers,



(a) Circular array.

(b) Linear array.

(c) Elliptical array.

Figure 6.18: Spatially averaged mean-squared error (6.29) as a function of temporal frequency. The plot marked with circles shows the error obtained with Tikhonov regularization for filter design, as in (6.23), while the plot marked with crosses shows the error obtained with the proposed regularization technique (6.40).

displaced to sample in an equiangular fashion an ellipse with semi-major and semi-minor axis equal to 1 m and 0.7 m, respectively. The desired field is that of a 2-D point source at $\rho_z = 2\,\mathrm{m}$, $\phi_z = \pi/3$.

It can be noticed that with the proposed regularization technique, the error is lower in the overall frequency band for all the loudspeakers configurations that have been considered. This is due to the exploitation of the statistics of the uncertainty in the filter design process, that adapts the regularization parameter to the actual perturbed model.

6.9 Main Results

- A technique for the synthesis of acoustic field is proposed. This technique is easy to adapt to arbitrary geometries of the loudspeaker array. Simulative results to validate the technique are presented.

- In order to overcome performance issue due to non-idealities in practical acoustic field synthesis systems, a technique to accurately render plane waves from arbitrary distribution of loudspeaker is proposed. The technique is designed to be robust to uncertainties in the loudspeaker propagation function.

Part III

Local Fourier Analysis, Synthesis and Processing

This page intetntionally left blank.

Chapter 7

Local Analysis and Processing of Acoustic Fields

Part II of this thesis showed ad hoc techniques for the analysis and synthesis of acoustic fields by means of arrays of microphones and loudspeakers, respectively. All the techniques that have been presented are close to the optimum in ideal scenarios, but when it comes to practical applications some of their limitations emerge. The major issues are: i) their poor robustness to real-world conditions (positioning errors, deviations from the ideal acoustic behavior, noise, etc.); ii) the difficulty in accomodating the modeling of the spatial evolution of an acoustic field (i.e. modeling of near field sources); and iii) their implementation complexity. Their implementation complexity, indeed, is still manageable with microphone and loudspeaker technologies commercially available today, where their per-unit cost limits the number of transducers to dozens or (in massive implementations) to few hundreds. Nowadays, the technological change from traditional to Micro Electro-Mechanical Systems (MEMS) microphones and loudspeakers paves the way to the availability of hundreds of acoustic transducers at an extremely low price. In this scenario the computational complexity of the techniques presented in Part II emerges as an actual limitation.

This chapter addresses the problem of analyzing and processing acoustic fields generated by sources in the proximity of the microphone array. In doing so, we rely on well established ideas borrowed from antenna array literature; those ideas and techniques are here extended and modified to suit acoustic applications. This approach presents many advantages, the most evident being the availability of a large *corpus* of widely studied algorithms; however, it should be remarked that these algorithms are designed based on assumptions that are not, in general, satisfied in the acoustic domain. In particular, many algorithms are designed based on far field assumptions; e.g., in the analysis case, sources are assumed to be in the far field, with wavefronts that impinge on the microphone array as planar wavefronts. Although this assumption is reasonable when the length of the array is really small compared to the wavelength, in most situations the proximity of acoustic sources implies a severe model violation, thus degrading the performance of the derived algorithm.

The adoption of well-established array processing techniques in the acoustic realm is enabled by the adoption of a novel analysis framework, based on the synergy between three elements:

1. the beam-based physical representation of acoustic field (introduced in Sec. 4.4), according to which acoustic fields can be expanded into a spectrum of beams propagating in different directions; this property is interpreted in the signal processing context as

2. the time-space redundant signal representation, based on the Local Fourier transform in Def. 8; and

3. the mapping of the transform coefficients onto a highly structured space with specific geometric properties.

The beam-based representation provides two main advantages over traditional representations. Indeed, unlike plane waves, beams are localized in space, meaning that only those beams passing near an observation point actually contribute to the field there. Moreover, beams intrinsically constitute a discrete set of basis functions, in the sense that beams originated at a discrete set of points and oriented to a discrete grid of directions can be used to represent an arbitrary acoustic field with a desired accuracy. Although beam-based representations are not popular, at our knowledge, among the acoustic signal processing community, they are gaining increasing interest in the antenna array processing community, e.g. in Maciel and Felsen [2002], Lugara et al. [2003], Shlivinski et al. [2004, 2005], Heilpern and Heyman [2013, 2014].

In this thesis, the beam-based representation is interpreted as a structured redundant frame-like representation, based on the Local Fourier transform (cfr. Def. 8). Frame-like ideas found a remarkable amount of applications in the overall signal processing community, among the others (without claims of completeness) in coding [Goyak et al., 1998, Burt and Adelson, 1983, Cvetković, 2003, Daubechies and DeVore, 2003, Benedetto et al., 2006], image denoising [Xu et al., 1994, Dragotti et al., 2003], transmission [Bernardini and Rinaldo, 2005, 2006a,b], signal reconstruction [Balan et al., 2006]. We refer the interested reader to Kovacevic and Chebira [2007] for a more comprehensive set of references. Redundant frame representations exhibit some great advantages in practical applications: mainly in terms of robustness to noise and to modeling mismatches. Moreover, in the specific case of acoustic signal processing, a frame-based analysis reveals important characteristics of the acoustic field. Indeed, in this chapter it is shown that frame-based analysis allows us to retrieve directional components of the acoustic field in the vicinity of an observation point.

This chapter hosts five main contributions. In Sec. 7.1, we present the plenacoustic representation that first appeared in Marković et al. [2013a]. We review the concept and the construction of the so-called *plenacoustic image*, which is mapped on a specific domain (the *ray space*). The ray space provides a suitable representation for the DoA of directional components of the acoustic field as the field itself varies over space. It is shown that in the ray space the acoustic field generated by acoustic primitives appear as linear patterns, thus enabling the easy extraction of their parameters.

The plenacoustic representation is exploited in Sec. 7.2 for the purpose of localizing acoustic sources placed in the vicinity of the microphone array. It is shown that acoustic sources can be localized by analyzing the energy distribution of acoustic field coefficients in the ray space. However, the realization propsed by Marković et al. [2013a] exhibits limitations due to poor resolution. In order to overcome these limitations, the high-resolution spectral analysis methodology presented in Sec. 5.2 is here adapted to the problem at hand. A simulative and experimental validation of this approach is provided.

The localization methodology presented in Sec. 7.2 can be employed as the fundamental building block for more complex acoustic field analysis methodologies. Sec. 7.3 presents a methodology for the extraction of a desired near-field sound source from the acoustic field captured by a microphone array, while attenuating interfering sources and noise. The methodology is developed in two steps: i) first, the acoustic sources are localized using the method presented in Sec. 7.2; ii) second, suitable spatial filters are designed to enhance/attenuate directional components of the acoustic field according to the geometrical

information provided by the first step, while attenuating also the noise component. The methodology presented in Sec. 7.3 is validated through simulations.

In Sec. 7.1, the plenacoustic representation is presented in the light of the local Fourier transform introduced in Sec. 2.5.1; we will show that the plenacoustic image can be conceived as the magnitude of the Local Fourier transform of the acoustic field captured by a linear microphone array, with the transform coefficients being mapped on the ray space.

In some applications the high redundancy of the plenacoustic representation is not needed, and a more parsimonious representation is desirable. Moreover, the plenacoustic representation emerging from Sec. 7.1 trns out to be non invertible. Section 7.5 shows that the theory of Gabor frames provides the foundations for a representation that is similar to the plenacoustic one (since they share the same domain, the ray space) but is less redundant and, furthermore, it features perfect reconstruction.

For the purpose of limiting the notational complexity without impairing the generality of the results, this chapter considers mainly a simplified scenario where a linear array of microphones deployed on the z axis is adopted. In this scenario, the signals acquired by the microphone array are independent on azimuth, hence all the acoustic quantities are considered just on the plane $y = 0$.

7.1 Plenacoustic Imaging

Plenacoustic imaging (also known as *soundfield imaging*) is a method for acoustic scene analysis originally presented in Marković et al. [2013a]. Authors in Marković et al. [2013a] adopt a linear array of microphones to capture the spatial information available on the acoustic scene in a local fashion. For this purpose, the authors measure the directional contributions to the observed acoustic field by using a linear array of microphones subdivided into overlapping sub-arrays, each composed by adjacent sensors. Data from each sub-array are used to estimate, through beamforming, the spatial spectrum [Stoica and Moses, 2004, p. 278], which represents an estimate of the acoustic energy carried by directional components measured at several points on the array. The magnitude of the spatial spectra is then represented as an image that is built by juxtaposition of the spatial spectra from individual sub-arrays. This image is referred to as *plenacoustic image* (or *soundfield image*).

The advantage of this acoustic scene analysis method is that acoustic primitives (as acoustic sources, reflectors, etc.) appear in the plenacoustic image as linear patterns. For example, a point source appears as a line in the plenacoustic image [Marković et al., 2013a, Fig. 3], while acoustic reflectors appear as wedges [Marković et al., 2013a, Fig. 4]. This property of the ray space makes it a favorable choice for many acoustic scene analysis applications, including the localization of multiple sources [Marković et al., 2013a], the estimation of the reflection coefficient [Marković et al., 2014] and others.

Review of the Ray Space. This paragraph briefly introduces the *ray space* as a convenient space where salient features of the plenacoustic representation of acoustic fields are apparent. Indeed, as described in Antonacci et al. [2008], acoustic primitives that generate the observed acoustic field are mapped in the ray space as linear patterns. In the following, a scenario where the acoustic field is observed only on a line is considered. In particular, the acoustic field is observed on the z axis, thus it is independent on the azimuth. Thanks to this consideration, without limiting the generality of the discussion, in the following the plane $y = 0$ is considered.

The ray space is introduced as the space encoding the parameters of acoustic rays. According to Marković et al. [2013a], an acoustic ray is the set of points (x, z) that satisfy the linear equation $z = \alpha x + \beta$, where the parameters (α, β) uniquely identify a ray in the plane $y = 0$, and they define the ray space.

The interested reader is referred to Antonacci et al. [2008] or to Marković et al. [2013a, Sec. II] for a formal definition of the ray space.

7.2 Source localization from Plenacoustic Images

The localization of acoustic sources in the vicinity of a microphone array is a typical application enabled by the plenacoustic representations. Using tools from the image pattern analysis literature, a plenacoustic image can be analyzed searching for linear patterns that identify acoustic primitives. In this context, one of the advantages of plenacoustic imaging is that several applicative problems can be solved using the same data, without resorting to ad-hoc representations suitable for the specific problem to be handled.

In Marković et al. [2013a] the authors employ a Minimum Variance Distortionless Response (MVDR) estimator (also known as *Capon* estimator) [Capon, 1969], [Stoica and Moses, 2004, Sec. 6.3.2] to compute the spatial spectrum from the data acquired by windowed portions of the overall array. This choice suffers from resolution issues. As a matter of fact, a trade-off issue arises between the width of the spatial window and the resolution of individual spatial spectra. Indeed, as the number of microphones insisting on a spatial window decreases, an higher number of windows is allowed (thus an higher number of spatial spectra composes the plenacoustic image), but, on the other hand, the resolution of each spatial spectrum decreases.

This trade-off has important consequences on applications of source localization. It is known from Marković et al. [2013a, Fig. 3] that a point source is mapped on the plenacoustic image as a line, and the estimate of the location can be accomplished by finding the parameters of linear patterns in the image. By decreasing W (the width of the spatial window), the location of the peaks in the plenacoustic image becomes undetermined due to the poor resolution of the image itself, but more data are available, as the plenacoustic image is composed by a higher number of spatial spectra.

This section proposes a method to overcome this trade-off using deconvolution techniques mutuated from research in aerospace acoustic imaging. Consider the case in which a single point source is present in the acoustic scene, and a continuous aperture is available to capture is acoustic field. In this ideal setting, spatial spectra appear as Dirac deltas in the direction of the acoustic source. In a real scenario, i.e. where the number of sensors is finite, a smearing of the spatial spectrum appears. Such smearing can be modeled by the Point Spread Function (PSF) [Ribeiro and Nascimento, 2011], a function that encodes the response of the array to a plane wave. Moreover, authors in Marković et al. [2013a] employ a rectangular window to partition the whole array into smaller sub-arrays. This choice maximizes the spatial resolution of the plenacoustic image, but introduces significant sidelobes in each spatial spectra, as the Fourier transform of the rectangular window is a cardinal sine function.

This section shows how to apply the deconvolution operator for improving the resolution of plenacoustic imaging. Moreover, in order to reduce the computational cost, a two-step version of the algorithm in Yardibi et al. [2008] is here presented, exploiting the information contained in the plenacoustic image. Indeed, the plenacoustic image is first analyzed to gather as much information as possible about the location of active acoustic sources. This

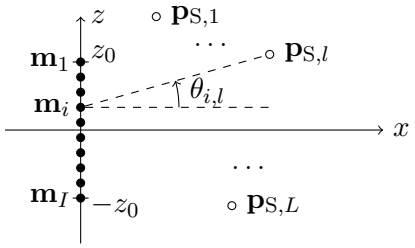


Figure 7.1: Source localization problem: microphone array and acoustic sources.

information is then used in a second stage to speed-up the deconvolution.

7.2.1 Notation

This paragraph introduces the notation used for the source localization application described in this section. Consider the setup in Fig. 7.1. L acoustic sources are located in $\mathbf{r}_{\mathrm{S},l} = [x_{\mathrm{S},l}, 0, z_{\mathrm{S},l}]^T$ and produce wide-band signals $s_l(t)$, $l = 1, \ldots, L$. An uniform linear array of I microphones is placed on the z axis between $z = -z_0$ and $z = +z_0$. The ith microphone is at $\mathbf{m}_i = [0, 0, q_0 - 2q_0 i/(I-1)]^T$, $i = 0, \ldots, I-1$. Spatial windowing is applied to the microphone array, using a rectangular window of width W. Thus, the array is partitioned into $M = I - W + 1$ sub-arrays. Microphones insisting on the window centered in $\mathbf{m}_i$ are denoted by $\mathbf{m}_j$, $j = i-(W-1)/2, \ldots, i+(W-1)/2$, and they acquire the signals $x_j(t)$. The signal $x_j(t)$ is processed with a filter bank to obtain $x_j(t, \omega_k)$, $k = 1, \ldots, K$, ω_k being the central frequency of the kth sub-band. These signals are then stacked into the vector $\mathbf{x}_i(t, \omega_k) = [x_{i-(W-1)/2}(t, \omega_k), \ldots, x_{i+(W-1)/2}(t, \omega_k)]^T$ and they can be expressed in matrix form as in (5.7)

$$\mathbf{x}_i(t, \omega_k) = \mathbf{A}_i(\omega_k)\mathbf{S}(t, \omega_k) + \mathbf{e}(t, \omega_k), \tag{7.1}$$

where $\mathbf{S}(t, \omega_k) = [s_1(t, \omega_k), \ldots, s_L(t, \omega_k)]^T$ contains the source signals and $\mathbf{e}(t, \omega_k)$ is a zero-mean additive noise with covariance matrix $\sigma^2\mathbf{I}$, $\mathbf{I}$ being the identity matrix of dimensions $W \times W$, and σ^2 being the power of the noise component. With the adoption of such a model, we assume that the noise component is uncorrelated between different microphones and its power is equal for each sensor. The matrix $\mathbf{A}_i(\omega_k)$ is the collection of sub-array steering vectors towards each of the L sources in the kth frequency sub-band, i.e.

$$\begin{aligned} \mathbf{A}_i &= [\mathbf{a}(\theta_{i,1}), \ldots, \mathbf{a}(\theta_{i,L})], \quad \mathbf{A}_i \in \mathbb{C}^{W \times L} \\ \mathbf{a}(\theta_{i,l}) &= [e^{j\left(i-\frac{W-1}{2}\right)\frac{\omega}{c}d\sin(\theta_{i,l})}, \ldots, e^{j\left(i+\frac{W-1}{2}\right)\frac{\omega}{c}d\sin(\theta_{i,l})}]^T, \end{aligned} \tag{7.2}$$

The symbol $\theta_{i,l}$ denotes the angle under which the lth source is seen from the microphone at $\mathbf{m}_i$. With reference to Fig. 7.1, it is

$$\theta_{i,l} = \arctan\left(\frac{z_{\mathrm{S},l} - z_0 + 2z_0 i/(I-1)}{x_{\mathrm{S},l}}\right). \tag{7.3}$$

7.2.2 Data Model

Following Marković et al. [2013a], and with reference to the notation introduced in the previous paragraph, a suitable beamforming technique computes the spatial spectrum $\tilde{p}_i(\theta, \omega_k)$ for the sub-array centered in $\mathbf{m}_i$ [Marković et al., 2013a], which can be considered as the estimate of the energy of acoustic rays reaching the central microphone at $\mathbf{m}_i$.

Spatial spectra $\tilde{p}_i(\theta, \omega_k)$ are then converted to $\tilde{p}_i(\alpha, \beta, \omega_k)$, whose domain is the ray space, through

$$\begin{aligned} \alpha &= \tan(\theta), \\ \beta &= z_0 - 2z_0 i/(I-1). \end{aligned} \tag{7.4}$$

The plenacoustic image $\mathcal{I}(\alpha, \beta, \omega_k)$ is, finally, the juxtaposition of $\tilde{p}_i(\alpha, \beta, \omega_k)$, i.e.

$$\mathcal{I}(\alpha, \beta_i, \omega_k) = \tilde{p}_i(\alpha, \beta, \omega_k), \tag{7.5}$$

where $\beta = \beta_i = z_0 - 2z_0 i/(I-1)$. In Marković et al. [2013a] authors also propose a wide-band version of plenacoustic imaging by averaging spatial spectra in a frequency sub-band fashion. In the following, it is specified if the discussion pertains narrow-band or wide-band signals. For compactness, however, the dependency on ω_k is omitted.

No matter what beamforming technique is used to extract $\tilde{p}_i(\theta)$, spatial spectra depend on the covariance matrix $\mathbf{R}_i$ of the signals acquired by the sensors under the ith spatial window. Upon assuming that noise and source signals are uncorrelated, the covariance matrix of array data can be modeled as

$$\mathbf{R}_i = \mathrm{E}\left\{\mathbf{x}_i(t)\mathbf{x}_i(t)^H\right\} = \mathbf{A}_i\mathbf{D}\mathbf{A}_i^H + \sigma^2\mathbf{I}, \tag{7.6}$$

where $\mathbf{D}$ is the covariance matrix of the L sources. For L uncorrelated sources with variance σ_l^2, the matrix $\mathbf{D}$ has the form

$$\mathbf{D} = \mathrm{diag}\left([\sigma_1^2, \ldots, \sigma_L^2]\right). \tag{7.7}$$

Denote by $\mathbf{h}_i(\theta)$ the spatial filter that performs the beamforming on the ith sub-array towards direction θ. Different techniques could be used to design such filter, such as Delay and Sum (DAS) beamforming [Stoica and Moses, 2004, Chap. 6] and MVDR. No matter what technique is used, the spatial spectrum is obtained as

$$\tilde{p}_i(\theta) = \mathbf{h}_i(\theta)^H\mathbf{R}_i\mathbf{h}_i(\theta). \tag{7.8}$$

For the DAS beamformer $\mathbf{h}_i(\theta) = \mathbf{a}(\theta)/W$ and

$$\tilde{p}_i(\theta) = \frac{1}{W^2}\left[\mathbf{a}(\theta)^H\mathbf{A}_i\mathbf{D}\mathbf{A}_i^H\mathbf{a}(\theta) + \sigma_2\mathbf{a}(\theta)^H\mathbf{a}(\theta)\right]. \tag{7.9}$$

In ideal conditions (i.e. infinite number of microphones) the first addend on the right-hand side of (7.9) is equal to zero for angles $\theta \neq \theta_{i,l}$ and equal to σ_l^2 for $\theta = \theta_{i,l}$. In a real situation, where the array resolution is finite, however, the spatial spectrum takes a non-zero value also for angles $\theta \neq \theta_{i,l}$. This effect is modeled in the literature through the PSF [Ribeiro and Nascimento, 2011]. These considerations can be applied also to beamforming techniques different from DAS.

Next paragraph summarizes a methodology, originally proposed in Yardibi et al. [2008], which attenuates the effect of the PSF through a deconvolution operation. The novel contribution of this section is in the application of such deconvolution methodology to the process of plenacoustic imaging.

7.2.3 Deconvolution

This paragraph reviews the deconvolution methodology, originally presented in Yardibi et al. [2008], for the estimation of sources location and power. This methodology is based on matching the sample estimate of the covariance matrix

$$\hat{\mathbf{R}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_i(t)\mathbf{x}_i(t)^H \tag{7.10}$$

to the model (7.6). It is worth noticing that the number L of sources is unknown in practical applications. As a consequence, a grid search over all the possible DoAs is performed. Let N be the number of possible directions of arrival of the source and let the plane wave components in the model (7.1) be computed on the grid of angles $\theta_{i,n}$, $n = 1, \dots, N$. Associated to the directions $\theta_{i,n}$ are the signals power σ_n^2. The matrix $\mathbf{D}$, introduced in (7.7), now generalizes to $\mathbf{D} = \text{diag}\left([\sigma_1^2, \dots, \sigma_N^2]\right)$. Under the assumption that $N \gg L$, $\mathbf{D}$ can be modeled as sparse; however, the overall methodology is robust to this assumption. The signal powers σ_n^2 and the additive noise power σ^2 are therefore estimated by solving the quadratic convex optimization problem

$$\begin{aligned} \underset{\sigma_n^2, n=1,\dots,N, \sigma^2}{\arg\min} \quad & \|\hat{\mathbf{R}}_i - \mathbf{A}_i \mathbf{D} \mathbf{A}_i^H - \sigma^2 \mathbf{I}\|_F^2 \\ \text{subject to} \quad & \sigma_n^2 \geq 0, \ n = 1, \dots, N, \\ & \sum_{n=1}^{N} \sigma_n^2 \leq \lambda_i, \ \sigma^2 \geq 0, \end{aligned} \tag{7.11}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Let $\hat{\mathbf{R}}_i = \hat{\mathbf{U}}_i \hat{\mathbf{\Gamma}}_i \hat{\mathbf{U}}_i^H$ be the eigen-decomposition of $\hat{\mathbf{R}}_i$. The parameter λ_i is given by

$$\lambda_i = \text{Tr}(\hat{\mathbf{\Gamma}}_i - \hat{\gamma}_i \mathbf{I}), \tag{7.12}$$

$\hat{\gamma}_i$ being the smallest eigenvalue of $\hat{\mathbf{R}}_i$. The number of unknowns in (7.11) is $N + 1$.

The problem formulation in (7.11) assumes the signal sources to be uncorrelated. In real acoustic scenes, however, acoustic sources can not be considered, in general, uncorrelated. This is, for instance, the case of non-point-like source, i.e. source whose spatial domain is extended. In this situation, there is a set of directions $\theta_{i,n}$ from which correlated signals impinge on the ith sub-array. In Yardibi et al. [2008] an extension to correlated sources is proposed. In this more general case, the matrix $\mathbf{D}$ is not diagonal any more, and the optimization problem can be reformulated as the semi-definite quadratic program

$$\begin{aligned} \underset{\sigma^2, \mathbf{D}}{\arg\min} \quad & \|\hat{\mathbf{R}}_i - \mathbf{A}_i \mathbf{D} \mathbf{A}_i^H - \sigma^2 \mathbf{I}\|_F^2 \\ \text{subject to} \quad & \mathbf{D} \geq 0, \ \text{Tr}(\mathbf{D}) \leq \lambda_i, \ \sigma^2 \geq 0. \end{aligned} \tag{7.13}$$

The number of unknowns is now $N^2 + 1$, therefore the solution of the problem requires a high computational cost. The increase in the computational cost becomes evident if we consider the case of wide-band sources, as the minimization must be accomplished K times, being K the total number of frequency bins of interest.

As shown in Marković et al. [2013a], it is possible to estimate the number of sources $\hat{L}$ along with their position $\hat{\mathbf{p}}_{\text{S},l}$ through pattern analysis on the plenacoustic image. This information is exploited to reduce the computational burden of the deconvolution. More specifically, the number of scanning directions N is replaced with the number of sources $\hat{L}$ estimated from a preliminary analysis of the non-deconvolved plenacoustic image, and considering the steering matrix

$$\mathbf{A}_i = [\mathbf{a}(\hat{\theta}_{i,1}), \dots, \mathbf{a}(\hat{\theta}_{i,\hat{L}})]. \tag{7.14}$$

The symbol $\hat{\theta}_{i,l}$ denotes the angle under which a source in $\hat{\mathbf{r}}_{\text{S},l}$ is seen from the microphone in $\mathbf{m}_i$. Thanks to this substitution, the number of unknowns in problems (7.11) and (7.13) reduces to $\hat{L} + 1$ and $\hat{L}^2 + 1$, respectively. However, the sparsity of the problem is not guaranteed any more. Moreover, an erroneous estimate $\hat{\mathbf{p}}_{\text{S},l}$ could lead to wrong results.

For the above reasons, acoustic sources are supposed not only to lie at the estimated position $\hat{\mathbf{p}}_{\mathrm{S},l}$, but also in their neighborhood, i.e. the deconvolution is performed on the grid of $\mathring{\mathbf{p}}_{\mathrm{S},l}$,

$$\begin{aligned}\mathring{\mathbf{p}}_{\mathrm{S},l} \in \{\mathbf{p}_{\mathrm{S},l} : \mathbf{p}_{\mathrm{S},l} = \hat{\mathbf{p}}_{\mathrm{S},l} + \boldsymbol{\delta}, \boldsymbol{\delta} = [\delta_x, \delta_y]^T, \\ -\delta_{\max} \leq \delta_x, \delta_y \leq \delta_{\max}\}.\end{aligned} \tag{7.15}$$

The steering vectors become

$$\mathbf{A}_i = [\mathbf{a}(\mathring{\theta}_{i,1}), \ldots, \mathbf{a}(\mathring{\theta}_{i,\hat{L}})], \tag{7.16}$$

where $\mathring{\theta}_{i,l}$ are the directions associated to $\mathring{\mathbf{p}}_{\mathrm{S},l}$ from the ith sub-array. The optimization problem is now sparse and is more robust against possible errors in the preliminary analysis. Finally, the deconvolved spatial spectra are

$$\tilde{p}_i(\theta_{i,n}) = \{\mathbf{D}\}_n^n, \tag{7.17}$$

where $\{\mathbf{D}\}_n^n$ denotes the nth entry on the diagonal of $\mathbf{D}$. The deconvolved plenacoustic image $\tilde{\mathcal{I}}(\alpha, \beta)$ becomes the juxtaposition of $\tilde{p}_{i,k}(\alpha, \beta)$, similarly to (7.5).

If the extension to wide-band signals is in order, one can proceed as in Azimi-Sadjadi et al. [2004] (cfr. Sec. 5.3) by averaging spatial spectra in a frequency sub-band fashion. In particular, the spatial spectral computed at all frequency bins in the frequency band of interest concur to the generation of a single plenacoustic image, with the product of their geometric and harmonic means

$$\tilde{p}_i(\theta_{i,n}) = \frac{l_2 - l_1}{\sum_{l=l_1}^{l_2} \tilde{p}_i(\theta_{i,n}, \omega_l)} \cdot \left(\prod_{l=l_1}^{l_2} \tilde{p}_i(\theta_{i,n}, \omega_l) \right)^{1/(l_2 - l_1)}, \tag{7.18}$$

where the integers l_1 and l_2 are the bin indexes corresponding to the cutoff frequencies of the frequency band of interest.

7.2.4 Localization Technique

Following Marković et al. [2013a], a two step localization methodology is adopted here, based on the analysis of the plenacoustic image. The first step consists in the identification and disambiguation of DoAs corresponding to the acoustic sources, which are estimated from the peaks of the computed spatial spectra. This operation is simplified by the inner nature of the ray space, which enables the clustering of DoAs corresponding to the same acoustic source on a linear pattern. The assignment of peaks to one of the sources is thus accomplished using a Hough transform [Duda and Hart, 1972].

The second step consists in finding an estimate of the location of the sources through a least-squares regression of the detected and clustered peak locations. In particular, the equation of a line in the ray space corresponding to the lth acoustic source is

$$\boldsymbol{\beta} = \boldsymbol{\alpha} x_{\mathrm{S},l} + z_{\mathrm{S},l}, \tag{7.19}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are two vectors collecting the estimated peak locations in the plenacoustic image. On rewriting (7.19) as

$$\boldsymbol{\beta} = \boldsymbol{\alpha}' \mathbf{r}_{\mathrm{S},l}, \tag{7.20}$$

with $\boldsymbol{\alpha}' = [\boldsymbol{\alpha}, \mathbf{1}]$ and $\mathbf{r}_{\mathrm{S},l} = [x_{\mathrm{S},l} z_{\mathrm{S},l}]^T$, one obtains that the line that best fits the estimated peaks in the least squares sense has parameters

$$\begin{bmatrix} x_{\mathrm{S},l} \\ z_{\mathrm{S},l} \end{bmatrix} = (\boldsymbol{\alpha}'^T \boldsymbol{\alpha})^{-1} \boldsymbol{a}'^T \boldsymbol{\beta}. \tag{7.21}$$

The line parameters in the rays space, i.e. $x_{\mathrm{S},l}$ and $z_{\mathrm{S},l}$ are the estimated position of the acoustic source in the Cartesian reference frame.

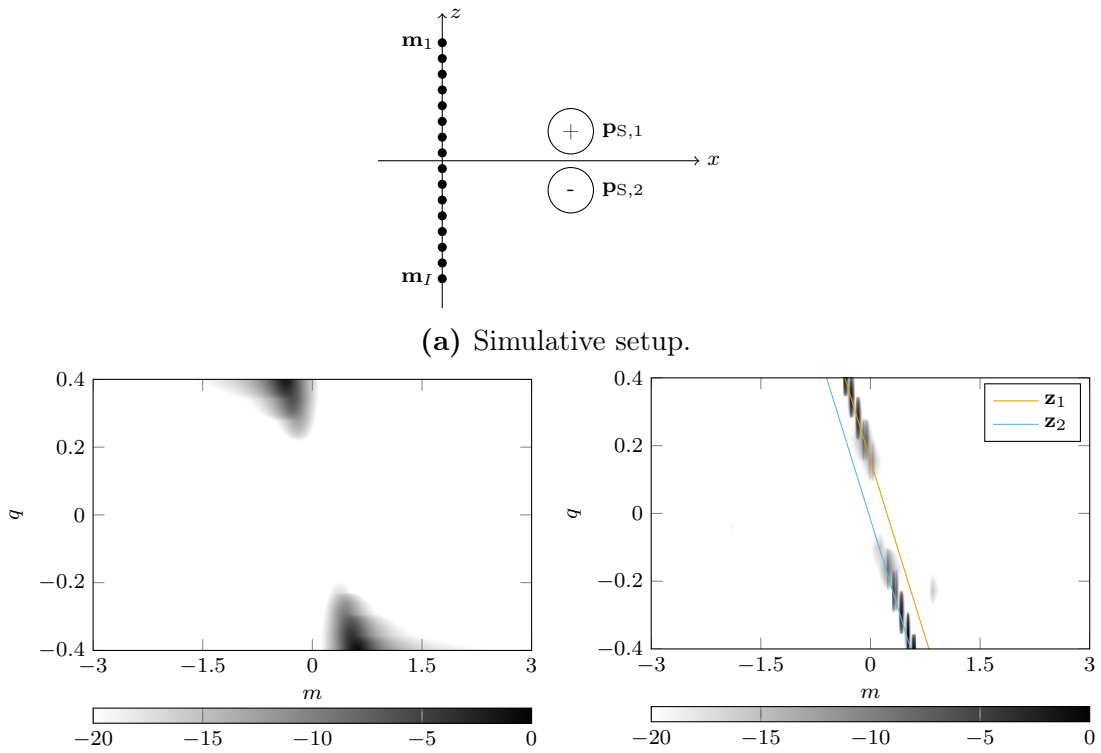
7.2.5 Results

This section presents simulations and experimental results to prove the advantages enabled by deconvolution in plenacoustic imaging. An uniform linear array of length 0.9 m ($q_0 = 0.45$ m) and composed by $I = 16$ microphones is considered. The array is divided into 14 sub-arrays, applying a rectangular spatial window of width $W = 3$. For all the simulations and the experiments, the signals acquired by the microphones are affected by an additive noise, whose power determines a Signal-to-Noise Ratio (SNR) equal to 10 dB. Non-deconvolved spatial spectra are obtained with MVDR beamformer, while deconvolved spatial spectra are obtained by solving the optimization problems in (7.11) or (7.13) with the CVX Matlab® toolbox [CVX Research, 2012, Grant and Boyd, 2008].

Extended Source. Consider the acoustic scene depicted in Fig. 7.2a. The acoustic source consists in two discs of radii $r = 0.08$ m and their centers are placed at $\mathbf{p}_{\mathrm{S},1} = [0.6\,\mathrm{m}, 0, 0.08\,\mathrm{m}]^T$ and $\mathbf{p}_{\mathrm{S},2} = [0.6\,\mathrm{m}, 0, -0.08\,\mathrm{m}]^T$. The signals emitted by the two discs are opposite in phase. Such setup has been simulated with the k-Wave toolbox [Treeby et al., 2012]. The plenacoustic image is expected to exhibit two parallel lines $\mathbf{z}_1$ and $\mathbf{z}_2$, corresponding to the centers of the discs $\mathbf{p}_{\mathrm{S},1}$ and $\mathbf{p}_{\mathrm{S},2}$, respectively. Moreover, the disc in $\mathbf{p}_{S,1}$ is expected to be occluded by the one in $\mathbf{p}_{S,2}$ for spatial windows in the half-space $z < 0$, and vice versa for spatial windows in the half-space $z > 0$. Finally, due to the mutual cancellation of contributions from the two discs, the image should exhibit a null region for spatial windows whose center is close to the origin. Fig. 7.2b and Fig. 7.2c show the plenacoustic images $\mathcal{I}(\alpha, \beta)$ and $\widetilde{\mathcal{I}}(\alpha, \beta)$, respectively, at $f = 1$ kHz. The non-deconvolved plenacoustic image $\mathcal{I}(\alpha, \beta)$ suffers from a smearing of the peaks and does not present significant energy in the range $-0.2 \leq \beta \leq 0.2$. On the other hand, the deconvolved plenacoustic image is composed by two parallel lines.

Estimation of the Radiation Pattern of Acoustic Sources. The increase of definition gained with the deconvolution enables further analysis on the plenacoustic image. As an example, Fig. 7.3 shows the actual radiation pattern $B(\theta)$ of the two extended sources considered above. $B(\theta)$ has been estimated by placing in the simulation setup 360 microphones around the source distribution. Fig. 7.3 also shows the patterns $\hat{B}(\theta)$ and $\tilde{B}(\theta)$ estimated from $\mathcal{I}(\alpha, \beta)$ and $\widetilde{\mathcal{I}}(\alpha, \beta)$, respectively. Note that $\hat{B}(\theta)$ takes values only at very few points, due to the absence of relevant peaks in the non-deconvolved spatial spectra composing $\mathcal{I}(\alpha, \beta)$. On the other hand, $\tilde{B}(\theta)$ closely approximates $B(\theta)$ for a significant number of points, thanks to the higher resolution of $\widetilde{\mathcal{I}}(\alpha, \beta)$.

Localization. Consider the setup depicted in Fig. 7.4a. Two point-like sources are located in $\mathbf{p}_{\mathrm{S},1} = [1\,\mathrm{m}, 0\,\mathrm{m} 0\,\mathrm{m}]^T$, and $\mathbf{p}_{\mathrm{S},2} = [(1 + \Delta x)\mathrm{m}, 0\,\mathrm{m}, 0\,\mathrm{m}]^T$, where Δx is in the range $(0.6\,\mathrm{m}, 1.6\,\mathrm{m})$. Localization is performed through suitable image pattern analysis on the plenacoustic images $\mathcal{I}(\alpha, \beta)$ and $\tilde{\mathcal{I}}(\alpha, \beta)$. The errors on the estimation of $\mathbf{p}_{\mathrm{S},1}$ and



(a) Simulative setup.

(b) $\mathcal{I}(\alpha, \beta)$.

(c) $\tilde{\mathcal{I}}(\alpha, \beta)$.

Figure 7.2: Comparison between non-deconvolved (Fig. 7.2b) and deconvolved (Fig. 7.2c) plenacoustic images for the setup in Fig. 7.2a. Grayscale plenacoustic images are represented in dB scale between 0 dB and −20 dB.

$\mathbf{p}_{S,2}$ from $\mathcal{I}(\alpha, \beta)$ are denoted as $\hat{e}_{S,1}$ and $\hat{e}_{S,2}$, respectively, while $\tilde{e}_{S,1}$ and $\tilde{e}_{S,2}$ are the localization errors from $\tilde{\mathcal{I}}(\alpha, \beta)$. One hundred independent realizations of the simulations are performed for each location $\mathbf{p}_{S,2}$.

Figure 7.4b shows the mean value of the localization errors as a function of the distance Δx. Notice that the deconvolution enables to accurately localize both sources in $\mathbf{p}_{S,1}$ and $\mathbf{p}_{S,2}$ starting from $\Delta x = 0.7\,\text{m}$, while without the deconvolution the localization is possible only from $\Delta x = 0.8\,\text{m}$.

Fig. 7.4c shows the average localization error for ten realizations of an experiment that reproduces the setup in Fig. 7.4a. Sources and microphones have been deployed in a modestly reverberant room. Notice that the localization error exhibits a trend similar to Fig. 7.4b, thus confirming the validity of the proposed technique also in real-word conditions. Moreover, with real-world data, the error $\hat{e}_{S,2}$ increases with Δx, while $\tilde{e}_{S,2}$ is almost constant.

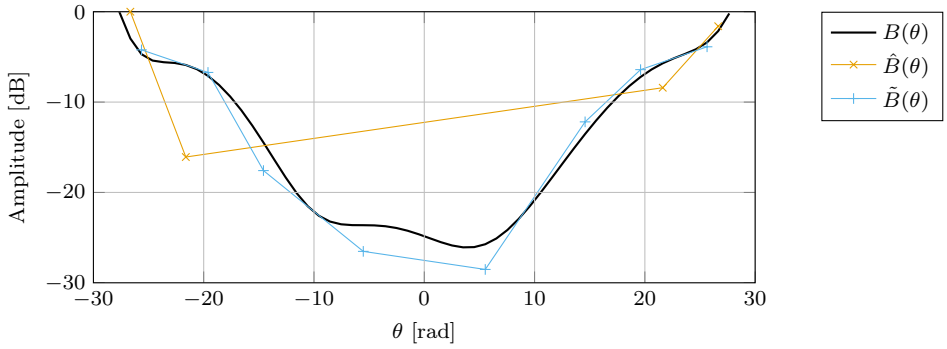
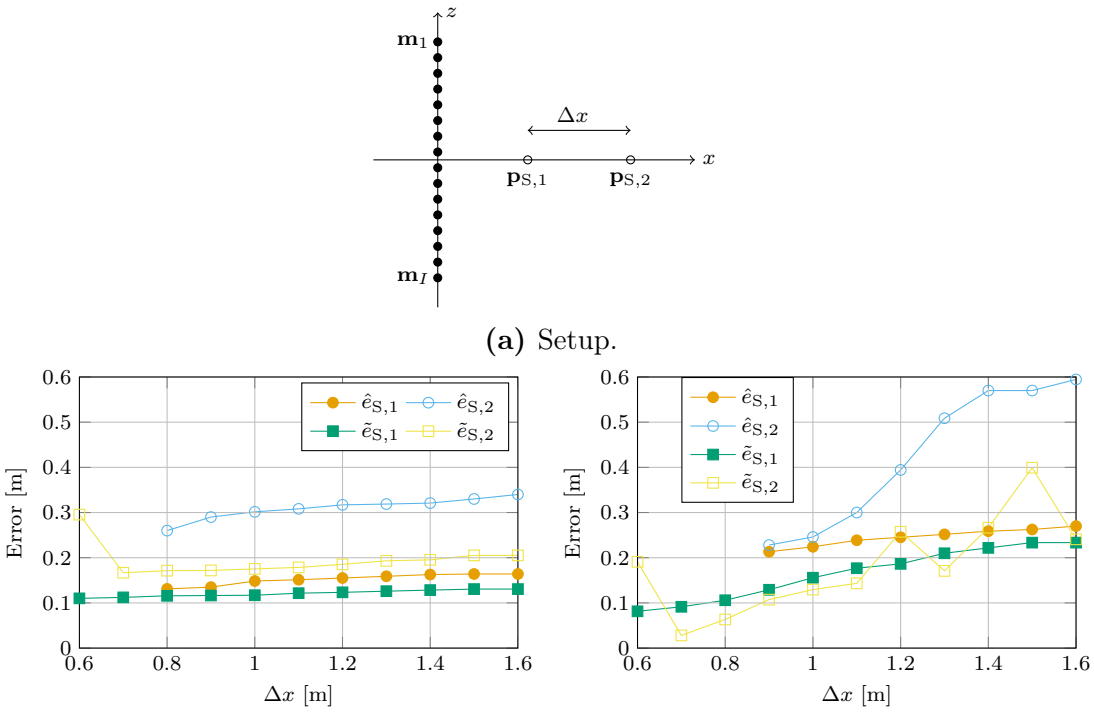


Figure 7.3: Comparison of radiation patterns estimated from $\mathcal{I}(\alpha, \beta)$ and $\tilde{\mathcal{I}}(\alpha, \beta)$.



(a) Setup.

(b) Simulations.

(c) Experiments.

Figure 7.4: Average localization error for the setup in Fig. 7.4a. Simulations (Fig. 7.4b) and experiments (Fig. 7.4c).

7.3 A Plenacoustic Approach to Source Separation

Source localization, as presented in Sec. 7.2, is a fundamental task both on its own and as a building block for more advanced applications. This section shows how to employ source localization and the plenacoustic imaging process in order to separate acoustic sources from interfering sources and noise.

Enhancement of a desired sound while attenuating interfering sources and background noise is a goal that has been pursued by the signal processing community for decades; the reader is referred to Benesty et al. [2005] for a comprehensive review of such research activity. An intuitive way of approaching the problem is to preliminarily gather spatial information about the acoustic scene, and then to focus on the source of interest through spatial filtering. Microphone arrays provide a technological mean to accurately capture spatial information [Benesty et al., 2008], with techniques borrowed from antenna array processing.

This approach presents some advantages, such as the availability of a large *corpus* of algorithms. However, it should be remarked that these algorithms are usually designed based on assumptions that are not, in general, satisfied in the acoustic domain: narrowband stationary signals, free-field propagation and sources in the far field with respect to the array. On the contrary, microphone array algorithms must be designed to deal with i) wideband time-varying signals; ii) reverberant environments; and iii) near-field sources.

The first issue is commonly solved by performing spatial filtering in a short-time-frequency domain [Allen, 1977, Benesty et al., 2012]. The second issue has been recently addressed, e.g. in Reuven et al. [2008], Markovich et al. [2009], Thiergart and Habets [2013]. In particular, authors in Thiergart and Habets [2013] propose a spatial filter that minimizes the diffuse plus noise power at filter output, subject to constraints on the desired spatial response. This way, the authors can specify a desired spatial response that enhances (attenuates) sounds coming from desired (undesired) source locations in the far-field; the constraints are set accordingly to an on-line estimate of the direction of arrival of the sources. However, this approach relies on the assumption of far-field sources impinging on the array as plane-wave wavefronts. This assumption can be violated in practical acoustic scenarios, thus degrading the performance of the spatial filter.

This section presents an approach for source separation that addresses the problem of working in near-field conditions, still using the simple and well established techniques borrowed from the array processing literature. A two step procedure is outlined, aimed at first localizing the source, and then using this information for separation purposes. As far as source localization is concerned, the approach presented in Sec. 7.2 is here adopted, and adapted to the case of possibly moving sources in non-anechoic environments.

As for the step of source separation, this section relies on the technique presented in Thiergart and Habets [2013], modified to work on a sub-array fashion. More specifically, a spatial filtering is performed for each spatially windowed sub-array, leaving undistorted the energy emitted by the desired sound source, while attenuating the interferer. Since sub-arrays are of smaller dimensions with respect to the overall array, the far-field propagation model is still applicable even in situations where the distance of the source is comparable with the overall array length. In this section, in line with the previous discussion, a rectangular spatial window is adopted.

Finally, a time reversal approach is used to apply a delay (possibly negative) and summing the desired signals at each sub-array, in such a way to propagate the signals back to the location of the desired source. In doing so, we rely on the source location estimated by inspection of the plenacoustic image.

7.3.1 Data Model

We consider the setup depicted in Fig. 7.1, with L (possibly moving) acoustic sources located at $\mathbf{p}_{\mathrm{S},l}(t) = [x_{\mathrm{S},l}(t), 0, z_{\mathrm{S},l}(t)]^T$, $l = 1, \ldots, L$, where t denotes time. The same microphone array setup described in Sec. 7.2.1 is adopted here. Let $x_i(t)$ denote the time-domain signal acquired by the ith microphone. In the following, a short-time Fourier analysis framework is considered (cfr. Def. 8), in which the signal acquired by the ith microphone is transformed in the signal $x_i(\tau, \omega)$ by

$$x_i(\tau, \omega) = \int_{-\infty}^{\infty} x_i(t) w(t - \tau) e^{-j\omega t}\, dt, \tag{7.22}$$

where $w(t)$ denotes a suitable real-valued temporal window function. For the purpose of clarity, in the following the dependency on time and frequency is omitted; it will be made again explicit Sec. 7.3.3, for the purpose of re-synthesizing time domain signals. Please notice that the proposed formulation accommodates the modeling of moving sources, as far as their position can be assumed to be fixed in the time interval covered by $w(t)$, as it is widely accepted in the literature.

Signals acquired by microphones are stacked into the vector

$$\mathbf{x}_i = [x_{i-(W-1)/2}, \ldots, x_{i+(W-1)/2}]^T. \tag{7.23}$$

Each acoustic source is assumed to be at a distance greater than the size of the sub-array, so that far-field propagation holds within sub-arrays. Under this assumption, the signal captured by the ith sub-array can be modeled as

$$\mathbf{x}_i = \mathbf{A}_i \mathbf{s}_i + \boldsymbol{\xi}_i + \mathbf{e}_i, \tag{7.24}$$

$\mathbf{s}_i = [s_{1,i}, \ldots, s_{L,i}]^T \in \mathbb{C}^{L \times 1}$ being a vector containing the L source signals. Let $\boldsymbol{\xi}_i \in \mathbb{C}^{W \times 1}$ denote the diffuse field impinging on the ith sub-array and $\mathbf{e}_i \in \mathbb{C}^{W \times 1}$ denote the additive microphone noise. $\mathbf{A}_i$ represents the collection of steering vectors [Stoica and Moses, 2004] towards each of the L sources

$$\begin{aligned} \mathbf{A}_i &= [\mathbf{a}(\theta_{i,1}), \ldots, \mathbf{a}(\theta_{i,L})], \quad \mathbf{A}_i \in \mathbb{C}^{W \times L} \\ \mathbf{a}(\theta_{i,l}) &= [e^{j\left(i - \frac{W-1}{2}\right)\frac{\omega}{c} d \sin(\theta_{i,l})}, \ldots, e^{j\left(i + \frac{W-1}{2}\right)\frac{\omega}{c} d \sin(\theta_{i,l})}]^T, \end{aligned} \tag{7.25}$$

where the symbol $\theta_{i,l}$ denotes the angle under which the lth sources is seen by the ith sub-array, i.e.

$$\theta_{i,l} = \arctan\left(\frac{y_{\mathrm{S},l} - q_0 + 2q_0(i-1)/(M-1)}{x_{\mathrm{S},l}}\right). \tag{7.26}$$

The covariance matrix of array data can be modeled as [Stoica and Moses, 2004]

$$\mathbf{R}_i = E[\mathbf{x}_i \mathbf{x}_i^T] = \mathbf{A}_i^H \boldsymbol{\Sigma}_i \mathbf{A}_i + \boldsymbol{\Xi}_i + \mathbf{E}_i, \tag{7.27}$$

where $\boldsymbol{\Sigma}_i = \mathrm{diag}(E[|s_{1,i}|^2], \ldots, E[|s_{L,i}|^2])$ contains the expected values of the squared amplitude of the L source signals on its diagonal, $(\cdot)^H$ denotes Hermitian transpose operator, $\boldsymbol{\Xi}_i = \sigma^2_{\mathrm{D},i}\boldsymbol{\Gamma}$ and $\mathbf{E}_i = \sigma^2_{\mathrm{N}}\mathbf{I}$. Let $\sigma^2_{\mathrm{D},i}$ and $\sigma^2_{\mathrm{N},i}$ denote the expected power of the diffuse field and of the microphone noise at the ith sub-array, respectively; $\boldsymbol{\Gamma}$ is the diffuse field coherence matrix [Cook et al., 1955] and $\mathbf{I}$ is the identity matrix.

The main goal of this section is to show how signal extraction of near-field sound sources can be conveniently accomplished in the plenacoustic framework. In particular, it is shown how data from individual sub-arrays concur in the estimation of the desired

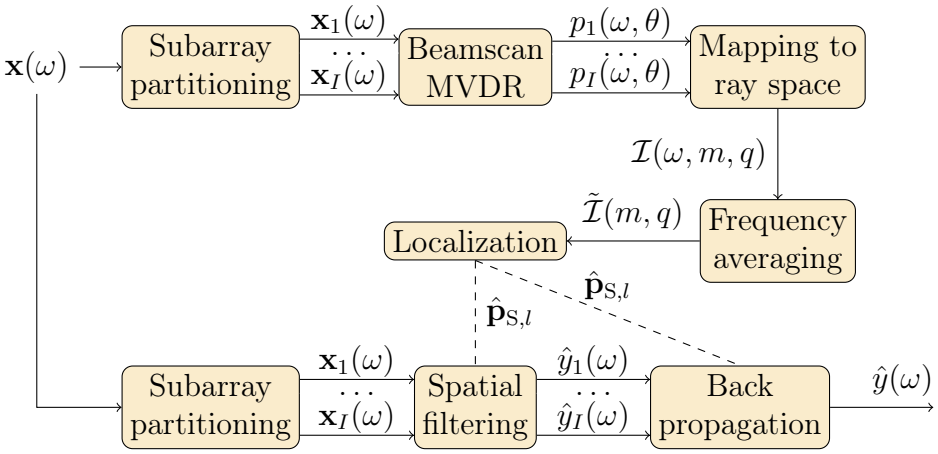


Figure 7.5: Block diagram for signal extraction of sources in the near-field.

signal through the application of a spatial filter informed on the location of the sound sources. The spatial filters are designed at sub-array level to attenuate or enhance sound coming from specific directions, derived from the source positions with basic geometric reasoning. Filter outputs from individual sub-arrays are then back-propagated from the acoustic center of each sub-array to the estimated source location of interest. Finally, signals are re-synthesized in the time domain. Figure 7.5 shows a block diagram of the overall system. Next paragraphs will detail all operations in Fig. 7.5.

7.3.2 Spatial Filtering

The goal of spatial filtering is to linearly combine the array data $\mathbf{x}$ such that sounds coming from different directions are enhanced or attenuated according to a desired directivity function, while attenuating both diffuse and noise field components. In this paragraph we focus on the whole array, since this is scenario considered in the original work [Thiergart and Habets, 2013], from which we derive the spatial filters used for our purposes.

The desired signal, i.e. a spatially filtered version of the source signals, can be written as

$$y = \mathbf{C}^T\mathbf{s}, \tag{7.28}$$

where $\mathbf{C} \in \mathbb{R}^{L\times 1}$ denotes the desired directivity function. A spatial filter $\mathbf{h}$ provides an estimate of y as a linear combination of array data

$$\hat{y} = \mathbf{h}^H\mathbf{x}, \quad \mathbf{h} \in \mathbb{C}^{W\times 1}. \tag{7.29}$$

In Thiergart and Habets [2013] the spatial filter is designed by minimizing the sum of diffuse and noise field powers at filter output, subject to the desired directivity function, i.e.

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \ \mathbf{h}^H(\mathbf{\Xi}+\mathbf{E})\mathbf{h} \quad \text{s.t} \ \ \mathbf{A}^H\mathbf{h} = \mathbf{c}. \tag{7.30}$$

By defining the Diffuse-to-Noise Ratio (DNR) $\sigma^2 = \sigma^2_{\mathrm{D}}/\sigma^2_{\mathrm{N}}$ and $\mathbf{J} = \sigma^2\mathbf{\Gamma} + \mathbf{I}$, the optimization problem in (7.30) can be rewritten in the equivalent form

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \ \mathbf{h}^H\mathbf{J}\mathbf{h} \quad \text{s.t} \ \ \mathbf{A}^H\mathbf{h} = \mathbf{c}, \tag{7.31}$$

whose solution is [Frost, 1972]

$$\hat{\mathbf{h}} = \mathbf{J}^{-1}\mathbf{A}\left(\mathbf{A}^H\mathbf{J}^{-1}\mathbf{A}\right)^{-1}\mathbf{c}. \tag{7.32}$$

7.3.3 Source Signal Extraction based on Plenacoustic Information

This paragraph shows how to extract the signal of a near-field sound source in the plenacoustic framework. Consider the partitioning of the overall microphone array into maximally overlapped spatially windowed sub-arrays, as described in Sec. 7.3.1. The spatial filtering technique introduced in Thiergart and Habets [2013] and reviewed in Sec. 7.3.2 is here adopted in order to enhance or attenuate sound coming from different directions with respect to the ith sub-array, while minimizing the diffuse and noise components. This filtering operation is supported by the information on the source locations, extracted from the plenacoustic image as described in Sec. 7.2.

Consider an acoustic scene in which L acoustic sources are present, whose locations have been estimated at $\{\hat{\mathbf{p}}_{\mathrm{S},l}\}$, using the method described in Sec. 7.2. A grid of directions $\{\hat{\theta}_{i,l}\}$ is set, and it is defined according to the estimated source positions $\{\hat{\mathbf{p}}_{\mathrm{S},l}\}$ as

$$\hat{\theta}_{i,l} = \arctan\left(\frac{\hat{y}_{\mathrm{S},l} - z_0 + 2z_0 i/(I-1)}{\hat{x}_{\mathrm{S},l}}\right). \tag{7.33}$$

Thanks to the use of the plenacoustic function, the accuracy of the DoA estimation at each sub-array is expected to be higher with respect to Thiergart and Habets [2013]. In fact, the DoA is extracted from the estimate $\{\hat{\mathbf{p}}_{\mathrm{S},l}\}$ of the source location, coming from the least squares line fitting described in Marković et al. [2013a], which enables to attenuate the impact of wrong DoA estimation on a small number of sub-arrays.

Let $\mathbf{A}_i$ denote the sub-array steering matrix computed on the set of directions $\{\hat{\theta}_{i,l}\}$. The desired directivity function for the ith sub-array is defined as

$$\{\mathbf{C}_i\}_l = \begin{cases} 1, & \text{if the } l\text{th source is desired, } (l = \check{l}) \\ 0, & \text{if the } l\text{th source is unwanted, } (l \neq \check{l}) \end{cases}, \tag{7.34}$$

where the symbol $\check{l}$ denotes the index of the desired source. The spatial filter $\hat{\mathbf{h}}_{i,\check{l}}$ to be applied to the ith sub-array for estimating the $\check{l}$th source is then computed through (7.32), and the corresponding filter output is given by

$$\hat{y}_{i,\check{l}} = \hat{\mathbf{h}}_{i,\check{l}}^H \mathbf{x}_i. \tag{7.35}$$

The signals $\hat{y}_{i\check{l}}$ estimated on a sub-array basis are then re-aligned and combined to form a single source signal estimate. For this purpose, a back-propagation approach is adopted, in which each sub-array signal $\hat{y}_{i,\check{l}}$ is back propagated from $\mathbf{m}_i$ to $\mathbf{p}_{\mathrm{S},\check{l}}$. The source signal estimate $\hat{y}_{\check{l}}$ is then obtained as the linear combination of the back-propagated sub-array signals

$$\hat{y}_{\check{l}} = 4\pi \sum_{i=1}^{I} \hat{y}_{i,\check{l}} \|\mathbf{p}_{\mathrm{S},\check{l}} - \mathbf{m}_i\| e^{i\frac{\omega}{c}\|\mathbf{p}_{\mathrm{S},\check{l}} - \mathbf{m}_i\|}, \tag{7.36}$$

where the back-propagation is performed assuming that the sound sources are isotropic point sources. We observe that this particular choice for the re-alignment of sub-array signals is arbitrary, as they could be re-aligned through back-propagation to an arbitrary point in space. However, the plenacoustic analysis framework provides an easy estimate of the actual source locations, hence it is natural to exploit this further information in the re-alignment phase.

Finally, the estimated source signal is re-synthesized in the time domain though an inverse Short-Time Fourier transform (cfr. Def. 8). In particular, denoting by $\hat{y}_{\check{l}}(\tau, \omega)$ the

signal estimated as in (7.36) from a frame windowed by $w(t-\tau)$ temporally centered in τ, at a temporal frequency ω, it results

$$\hat{y}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{y}_{\tilde{l}}(\tau, \omega) w(t-\tau) e^{j\omega t} \, d\omega \, d\tau. \tag{7.37}$$

7.3.4 Results

This section presents some simulations aimed at validating the proposed approach to acoustic signal extraction. All the simulations are performed in anechoic conditions with $I = 16$ microphones with spacing $d = 6\,\text{cm}$. The width of the spatial window determining the sub-arrays is varied among simulations and will be specified for each case. The sampling frequency is set to 44.100 kHz. The short-time analysis is accomplished using a Hanning window of length 23.21 ms, 50% overlap. The speed of sound is fixed to $340\,\text{m}\,\text{s}^{-1}$. The set of directions $\{\theta_{i,l}\}$ is obtained by uniformly sampling the interval $[-\pi/2, \pi/2]$ in $M = 65$ points. The number of the acoustic sources is fixed to $L = 2$. The source signals are speech signals from Eur [2008, Tracks 49 (female) and 50 (male)]. In the following, the female speech is assigned to the source with index 1, while the male speech is assigned to the source with index 2. Prior to the processing, the source signals are filtered with a bandpass filter whose cutoff frequencies are 500 Hz and 5 kHz. In all the simulations, the power of the diffuse noise field is set to $\sigma_{\text{D}}^2 = 0$. In the second and third simulations the variance of the additive noise σ_{N}^2 is set so that the signal-to-noise ratio is 20dB.

The performance of the signal extraction approach is evaluated in terms of the Signal-to-Interference Ratio (SIR), defined as the ratio between the energy of the desired source signal and the sum of noise and undesired sources energy [Vincent et al., 2006]

$$\text{SIR} = 10 \log_{10} \frac{\|\hat{y}_{\tilde{l}}(t)\|^2}{\|\epsilon(t)\|^2}, \tag{7.38}$$

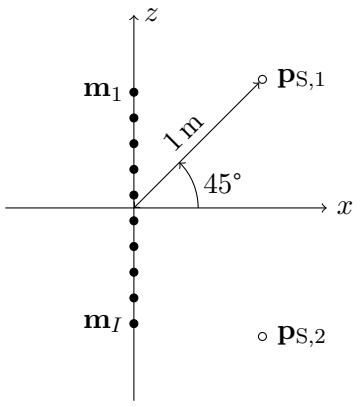
where $\epsilon(t)$ is defined as the difference of two contributions:

1. the orthogonal projection of the estimated signal $y_{\tilde{l}}(t)$ onto the space spanned by all source signals (i.e. $s_l(t)$, $l = 1, \ldots, L$); and
2. the orthogonal projection of $y_{\tilde{l}}(t)$ onto the space spanned by the original desired source signal $s_{\tilde{l}}(t)$.

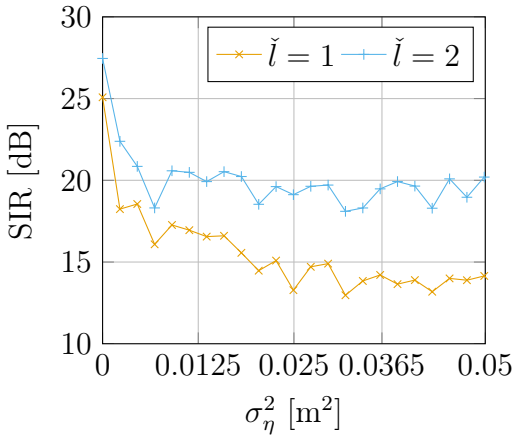
The SIR metric has been computed using the Matlab implementation provided in Vincent et al. [2012].

The first simulation evaluates the impact of the localization error on the overall performance, according to the setup depicted in Fig. 7.6a. For this purpose, an additive random variable to the true source locations is introduced, such that $\hat{\mathbf{p}}_{\text{S},l} = \mathbf{p}_{\text{S},l} + \eta$, η being a bi-variate Gaussian random variable with covariance matrix $\sigma_\eta^2 \mathbf{I}$ and zero mean. The sources are placed at $\mathbf{p}_{\text{S},1} = [\sqrt{2}/2\text{m}, 0\,\text{m}, \sqrt{2}/2\text{m}]^T$ and $\mathbf{p}_{\text{S},2} = [\sqrt{2}/2\text{m}, 0\,\text{m}, -\sqrt{2}/2\text{m}]^T$. In order to assess only the impact of the localization error, in this simulation the additive noise power is set to $\sigma_{\text{N}}^2 = 0$. The standard deviation of the localization error σ_η is let vary between 0 m and 0.23 m. For each value of σ_η, 50 realizations of η are simulated. Figure 7.6b shows the SIR averaged over all the realizations for each value of σ_η^2.

We observe that the localization error introduces some impairments in the source extraction system, but even in cases where the localization error is high, there is still an acceptable degree of separation of the two sources. We observe that introducing a localization error with variance $\sigma_\eta^2 > 0.01$, the SIR for both sources remains approximately constant.

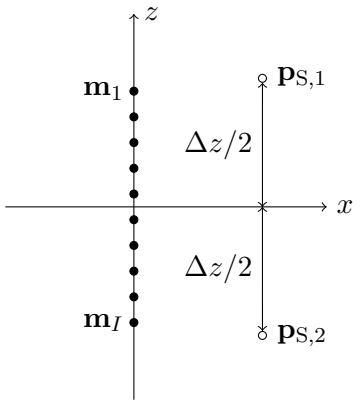


(a) Simulation 1.

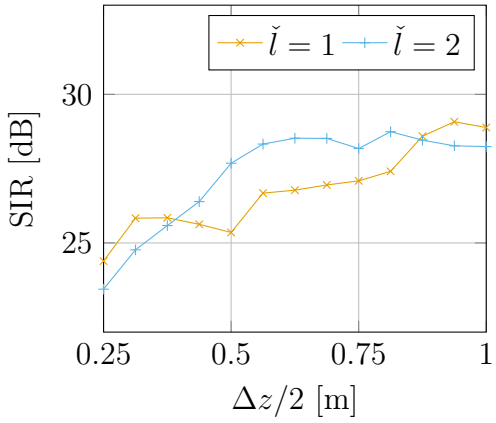


(b) Results from simulation 1.

Figure 7.6: Impact of the localization error on SIR.



(a) Simulation 2.

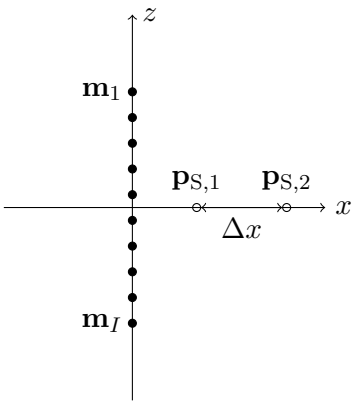


(b) Results from simulation 2.

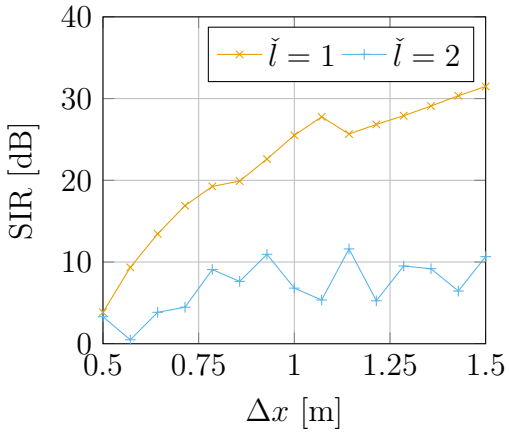
Figure 7.7: Impact of the transversal source distance Δz on SIR.

The second simulation is aimed at assessing the performance of the source extraction approach when two sources are placed at $\mathbf{p}_{S,1} = [1, 0, \Delta z/2]^T$ and $\mathbf{p}_{S,2} = [1, 0, -\Delta z/2]^T$. Figure 7.7a shows the simulative setup. The proposed method is still able to extract the two sources with SIR $\approx$ 25dB even when $\Delta z = 50\,\text{cm}$, which in the considered setup is equivalent to an angular separation of 25°. The extraction performance is further improved when Δz is increased.

Finally, the third simulation is aimed at assessing the performance of the source extraction approach when two sources are aligned on the x axis, placed at $\mathbf{p}_{S,1} = [0.5\,\text{m}, 0, 0\,\text{m}]^T$ and $\mathbf{p}_{S,2} = [(0.5 + \Delta x)\text{m}, 0, 0\,\text{m}]^T$. Figure 7.8a shows the simulative setup. The proposed method is able to extract the two sources even in this challenging scenario. In particular, the sound source closer to the microphone array is extracted with a SIR $>$ 20dB when $\Delta x > 0.8\,\text{m}$. The sound source farthest from the microphone array is still extracted but with a lower SIR.



(a) Simulation 3.



(b) Results from simulation 3.

Figure 7.8: Impact of the co-linear source distance Δx on SIR.

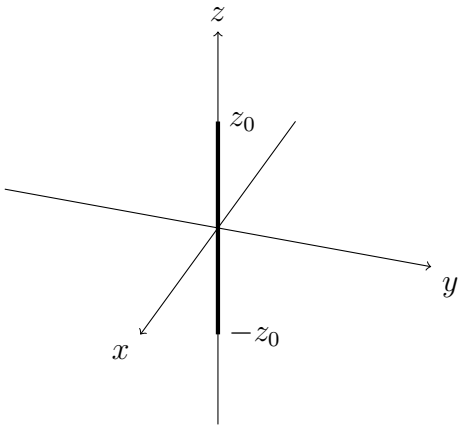


Figure 7.9: Linear aperture on the z axis.

7.4 Plenacoustic Imaging as a Local Fourier transform

This section reinterprets the plenacoustic representation discussed in previous sections in the light of the local Fourier transform (cfr. Def. 8).

7.4.1 Ideal Plenacoustic Imaging

In order to introduce the concepts in a progressive fashion, consider first the case in which a pressure sensitive aperture is deployed on the z axis, between $z = -z_0$ and $z = z_0$, as depicted in Fig. 7.9. Notice that in this scenario the signal received by the microphone array is independent on azimuth, hence any acoustic system built according to the configuration in Fig. 7.9 is not able to discriminate the azimuthal direction. The pressure sensitive aperture enables to capture the acoustic field $P(\mathbf{r}, \omega)$ on the z axis between $z = -z_0$ and $z = z_0$, i.e. $\mathbf{r} = [0, 0, z]^T$, with $-z_0 \leq z \leq z_0$; thus, the captured acoustic field depends only on z and on the temporal frequency ω, hence it can be written as $P(z, \omega)$. In order to simplify the notation, in the following we omit the dependency on ω.

Consider a window function $\psi(z)$ that allows us to select portions of the z axis. Plenacoustic imaging, as it has been introduced in previous sections prescribes the adoption of a rectangular window. The local Fourier transform (2.74) is obtained by multiplying the acoustic field $P(z)$ with shifted copies of the window function $\psi(z)$, i.e.

$$\left[\tilde{\mathbf{P}}\right]_{m,n}(\omega) = \int_{-z_0}^{z_0} P(z,\omega)\psi^*(z - m\bar{z})e^{-j\frac{\omega}{c}n\bar{k}_z(z-m\bar{z})}\,dz, \tag{7.39}$$

where $m = 0, \ldots, M-1$ is an integer representing the spatial displacement along the z axis and $n = 0, \ldots, N-1$ is an index representing the spatial frequency. The parameters $\bar{z}$ and $\bar{k}_z$ denote the sampling intervals on the z axis and on the spatial frequency axis, respectively.

For convenience, as will be clear in the following, the local Fourier transform coefficients have been arranged in the matrix $\tilde{\mathbf{P}}$. As it has been shown in Sec. 4.1.1, the complex exponential term in (7.39) can be interpreted as a plane wave with DoA θ_n given by $\cos(\theta_n) = n\bar{k}_z$. Equation (7.39) is interpreted as a set of beamforming operations each insisting on a portion of the overall aperture centered at $z - m\bar{z}$ and steered towards direction $\theta_n = \arccos(n\bar{k}_z)$. Notice that this interpretation coincides with the notion of *ideal soundfield camera* in Marković et al. [2013a, Sec. III-A], where the window function $\varphi(z)$ is chosen to be a rectangular window.

Following Marković et al. [2013a, Sec. III-A], the energy of individual local Fourier transform coefficients $|\tilde{\mathbf{P}}_{m,n}|^2$ is mapped onto the ray space by

$$\begin{aligned} \beta &= m\bar{z}, \\ \alpha &= \arctan(\arccos(n\bar{k}_z)) = \frac{\sqrt{1-(n\bar{k}_z)^2}}{n\bar{k}_z}, \end{aligned} \tag{7.40}$$

where β represents the z coordinate of the beam origin and α is the angular coefficient of the line coincident with the beam axis.

7.4.2 Plenacoustic Imaging with a Microphone Array

The previous paragraph showed the idea of plenacoustic imaging in an ideal scenario, where a pressure sensitive aperture was available. This paragraph shows how a plenacoustic image can be acquired by means of a microphone array that samples the continuous aperture in Fig. 7.9. Indeed, consider a linear array of microphones deployed on the z axis; microphones sample a portion of the z axis of length $2z_0$, ranging from $z = -z_0$ to $z = z_0$; Fig. 7.10 shows the setup.

The local Fourier transform coefficients in (7.39) can be approximated by applying a quadrature rule to the integral over the z axis, where quadrature points are the location of the microphones in the array. This approximation yields

$$\left[\tilde{\mathbf{P}}\right]_{m,n}(\omega) \approx d\sum_{i=0}^{I} P(z_i,\omega)\varphi^*(z_i - m\bar{z})e^{-j\frac{\omega}{c}n\bar{k}_z(z_i-m\bar{z})}, \tag{7.41}$$

where I is the total number of microphones and $d = 2z_0/(I-1)$ is the distance between adjacent microphones. Figure 7.11 shows an illustrative plenacoustic image generated by two acoustic sources; the lines $\mathcal{R}_{\mathbf{r}_1}$ and $\mathcal{R}_{\mathbf{r}_1}$ denote the linear patterns corresponding to the two acoustic sources.

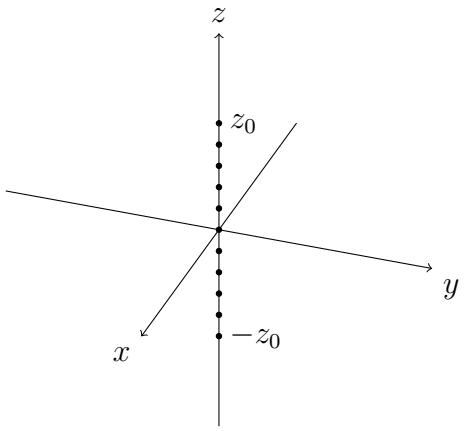


Figure 7.10: Linear array on the z axis.

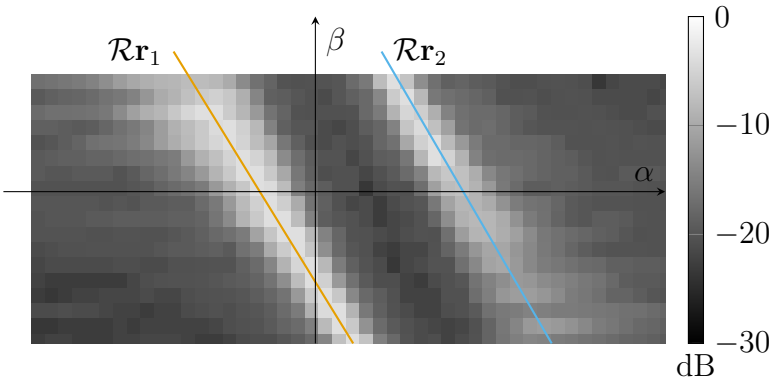


Figure 7.11: Illustrative plenacoustic image.

7.4.3 Interpretation

Consider a single spatial window in (7.41), i.e. fix m. The mth row of the matrix $\tilde{\mathbf{P}}(\omega)$ is interpreted as the spatial spectrum computed from the array data multiplied by the spatial window function centered in $\mathbf{m}_i = [0, 0, z - m\bar{z}]^T$, as shown in Fig. 7.12 for the case of a rectangular window. Indeed, by fixing m and adopting a rectangular spatial window of width W, (7.41) becomes

$$\tilde{P}_n(\omega) \approx \Delta_z \sum_{i=m-(W-1)/2}^{m+(W-1)/2} P(z_i, \omega) e^{-j\frac{\omega}{c} n \bar{k}_z (z_i - m\bar{z})}, \tag{7.42}$$

which is readily interpreted as a DAS beamforming operation [Stoica and Moses, 2004, Sec. 6.3.1] on a partial set of array data. In the nomenclature introduced in previous sections, this partial set is referred to as a *sub-array*. In the light of this interpretation, (7.41) is considered as the collection of multiple DAS beamforming operation, each computed from a specific portion of array data.

The spatial spectral computed from the DAS beamforming estimator are known to exhibit the same resolution issues of the periodogram spectral estimator [Stoica and Moses, 2004, Chap. 2]. In fact, it is shown in [Stoica and Moses, 2004, p. 278] that the beamwidth of the spatial filter used to compute the DAS beamforming estimator is approximately proportional to the inverse of the sub-array length. Adopting sub-arrays of reduced length with respect to the overall array yields to resolution issues, as it has been shown in Sec. 7.2.

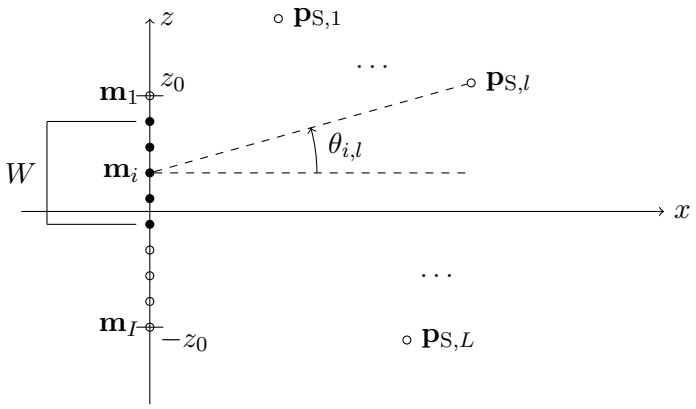


Figure 7.12: Interpretation of the plenacoustic imaging process.

7.4.4 Redundancy of the Plenacoustic Representation

The acoustic field in the ray space is represented by MN coefficients; the number of steering directions N is usually chosen to match the desired resolution and can be in the order of hundreds; on the other hand, the number of window centers I is determined by the specific choice for the window function and by the overlap between sub-arrays. Upon adopting the rectangular window of width W and setting the maximum overlap between windows, M is determined as

$$M = I - W + 1. \tag{7.43}$$

Typical array configurations, realized with traditional microphone technologies (and not with MEMS), consist of tenth of microphones; thus, the acoustic field is represented in the acoustic domain by a number of coefficients in the order of thousands.

Just to make an illustrative example, consider a microphone array composed by $I = 32$ microphones and partitioned using maximally overlapped rectangular windows of width $W = 5$; this yields to $M = 28$ window centers. Setting the desired resolution of the acoustic image in the DoA dimension to 5° yields $N = 180°/5° = 72$. Thus the total number of coefficients in the plenacoustic representation is $MN = 2016$. This number has to be compared with the original amount of coefficients needed to represent the acoustic field captured by the microphone array, which is given by the number of microphones $I = 32$ itself. Thus, the redundancy of this specific illustrative plenacoustic representation is $2016/32 = 63$.

The high redundancy intrinsic to the plenacoustic representation enables two major benefits over non-redundant representations:

- the information contained in the signal captured by the microphone array is cast in a form where its salient characteristics are immediately understandable; this form, as it has been detailed in past sections, is provided by the ray space;
- redundancy attenuates the effect of noise in the array data; this feature is fundamental in many applications; indeed, it has been exploited for localization and source extraction purposes.

7.5 The Plenacoustic Transform

The plenacoustic representation discussed in Chap. 7 exhibits many interesting properties that make it a favorable choice for many applications. However, there are some applications

where its properties are not optimal. The major of its pitfalls is that the plenacoustic representation does not have the concept of a full-fledged signal transformation embedded in it. As a matter of facts, no tools have been developed to invert the plenacoustic representation, i.e. to compute the acoustic field given its plenacoustic representation.

Another restriction set by the plenacoustic approach has it has been presented in Chap. 7 is the it constrains us to use rectangular windows to partition the whole array into smaller portions. Rectangular windows used in the context of the plenacoustic representation exhibits the same characteristics widely studied in the field of time-frequency analysis of signals. Precisely, if we narrow the window (i.e. we consider smaller sub-arrays) the spatial resolution increases, meaning that we are provided with more analysis points along the array; however, since the Fourier transform of the rectangular window is the cardinal sine function, it is easy to notice that the directional resolution worsens. Hence, the width of the rectangular window function balances the spatial and directional resolutions. This consideration motivates the impossibility of having a plenacoustic image with both good spatial and directional resolutions; a technique to overcome such limitation using rectangular windows and performing a directional analysis of data based on the sub-array covariance matrix has been presented in Sec. 7.2.

A different choice for the window function paves the way to many interesting properties. In this chapter we consider the Gaussian window. This choice is supported mainly by three motivations:

1. the Gaussian window exhibits the best trade-off between resolution in spatial and directional domains;

2. the Gaussian window enables a strong physical interpretation, in the sense that it allows us to build a frame-like representation whose constituting wave objects are acoustic beams (cfr. Sec. 3.8 and Sec. 4.4);

3. the Gaussian window makes it possible to define an invertible transformation that allows us to represent acoustic signals in the ray space.

The last motivation paves the way to the definition of a new full-fledged transformation for dealing with acoustic signals in the ray space. As it was claimed in Chap. 1 this is the final goal of this thesis. In the following of this section we define the *plenacoustic transform* starting from the setting of discrete Gabor frames [Qiu and Feichtinger, 1995, Werther et al., 2005].

7.5.1 The Plenacoustic Transform as a Linear Operator

Consider a family of discrete length-L functions $\psi_{m,n} \in \mathbb{C}^L$, obtained as modulation and translation of a prototype function $\psi(l)$, i.e.

$$\psi_{m,n}(l) = \psi(l - na)e^{j2\pi m(l-na)/M} = \mathcal{M}_{m/M}\mathcal{T}_{na}\psi(l), \tag{7.44}$$

where $m = 0, \dots, M-1$, $n = 0, \dots, N-1$ and $l = 0, \dots, L-1$ are integer indexes; the parameters a, b, M, N and L are integers that assume the following meanings:

- a is the translation step of the windows on the z axis;
- b is the modulation step of the windows on the spatial frequency domain;
- M is the total number of spatial frequencies;
- N is the total number of spatial translations;

- L is the number of microphones in the whole array, i.e. the number of samples of the function $P(z)$ as observed by the microphone array.

The parameters that have been just introduced must satisfy

$$L = aN = bM. \tag{7.45}$$

In (7.44) the symbols $\mathcal{T}$ and $\mathcal{M}$ are used to denote the translation and modulation operators, defined as

$$\mathcal{T}_\zeta P(z) = P((z-\zeta) \mod L), \tag{7.46}$$
$$\mathcal{M}_\omega P(z) = P(z)e^{j2\pi\omega z}, \tag{7.47}$$

respectively. The family of functions satisfying (7.44) is the *Gabor system* (ψ, a, M).

In this paragraph we define an algebraic setting where the plenacoustic transform, in both its analysis and synthesis operations, appears as a linear transformations of array data. In the derivation of the intended setting we follow the literature on discrete Gabor frames, e.g. Qiu and Feichtinger [1995], Strohmer [1998], Feichtinger et al. [2009], Moreno-Picot et al. [2010].

Let us define the matrix $\mathbf{\Psi} \in \mathbb{C}^{L\times MN}$ that contains the window $\psi_{m,n}$ as its $(m + nM + 1)$th column, i.e.

$$\begin{aligned}\mathbf{\Psi} &= \begin{bmatrix} \psi_{0,0}(0) & \psi_{1,0}(0) & \dots & \psi_{0,1}(0) & \dots & \psi_{M,N}(0) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \psi_{0,0}(L-1) & \psi_{1,0}(L-1) & \dots & \psi_{0,1}(L-1) & \dots & \psi_{M,N}(L-1) \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\psi}_{0,0} & \boldsymbol{\psi}_{1,0} & \dots & \boldsymbol{\psi}_{M,N} \end{bmatrix},\end{aligned} \tag{7.48}$$

where the second equality has been derived by denoting $\boldsymbol{\psi}_{m,n} = [\psi_{m,n}(0), \dots, \psi_{m,n}(L-1)]^T$.

Let us denote the acoustic signal captured by the microphone array as the vector $\mathbf{p} \in \mathbb{C}^L$. We define the matrix $\tilde{\mathbf{P}} \in \mathbb{C}^{M\times N}$ containing the plenacoustic coefficients of the acoustic field

$$\tilde{\mathbf{P}} = \begin{bmatrix} \langle \mathbf{p}, \boldsymbol{\psi}_{0,0} \rangle & \langle \mathbf{p}, \boldsymbol{\psi}_{0,1} \rangle & \dots & \langle \mathbf{p}, \boldsymbol{\psi}_{0,N-1} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{p}, \boldsymbol{\psi}_{M-1,0} \rangle & \langle \mathbf{p}, \boldsymbol{\psi}_{M-1,1} \rangle & \dots & \langle \mathbf{p}, \boldsymbol{\psi}_{M-1,N-1} \rangle \end{bmatrix}. \tag{7.49}$$

The vector $\tilde{\mathbf{p}}$, defined from $\tilde{\mathbf{P}}$ by stacking its columns so that the (m, n)th element of $\tilde{\mathbf{P}}$ corresponds to the $(m + nM + 1)$th element of $\tilde{\mathbf{p}}$, can be obtained as a linear transformation of array data, as in Moreno-Picot et al. [2010, Eq. 2]

$$\tilde{\mathbf{p}} = \mathbf{\Psi}^H \mathbf{p}. \tag{7.50}$$

With the notation just introduced, the frame operator (cfr. Sec. (2.4)) is defined as the matrix [Moreno-Picot et al., 2010, Eq. 5]

$$\mathbf{S} = \mathbf{\Psi}\mathbf{\Psi}^H. \tag{7.51}$$

If the matrix $\mathbf{\Psi}$ has rank L, the Gabor system (ψ, a, M) represents a frame in $\mathbb{C}^L$, as demonstrated in Strohmer [1998]; in this case, it is possible to reconstruct $\mathbf{p}$ from its

plenacoustic coefficients $\tilde{\mathbf{P}}$ by using a dual Gabor system composed by the family of function $\tilde{\psi}$. The reconstruction formula, in this case, reads

$$\mathbf{p} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle \mathbf{p}, \tilde{\boldsymbol{\psi}}_{m,n} \rangle \boldsymbol{\psi}_{m,n}, \tag{7.52}$$

where the elements $\{\tilde{\boldsymbol{\psi}}_{m,n}\}$ are defined similarly to (7.44).

We define now $\tilde{\boldsymbol{\Psi}} \in \mathbb{C}^{L \times MN}$ as the matrix having $\tilde{\boldsymbol{\psi}}_{m,n}$ as its $(m+nM+1)$th column, i.e.

$$\begin{aligned} \tilde{\boldsymbol{\Psi}} &= \begin{bmatrix} \tilde{\psi}_{0,0}(0) & \tilde{\psi}_{1,0}(0) & \dots & \tilde{\psi}_{0,1}(0) & \dots & \tilde{\psi}_{M,N}(0) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\psi}_{0,0}(L-1) & \tilde{\psi}_{1,0}(L-1) & \dots & \tilde{\psi}_{0,1}(L-1) & \dots & \tilde{\psi}_{M,N}(L-1) \end{bmatrix} \\ &= \begin{bmatrix} \tilde{\boldsymbol{\psi}}_{0,0} & \tilde{\boldsymbol{\psi}}_{1,0} & \dots & \tilde{\boldsymbol{\psi}}_{M,N} \end{bmatrix}. \end{aligned} \tag{7.53}$$

If $\tilde{\boldsymbol{\Psi}}$ is a right inverse for $\boldsymbol{\Psi}$, i.e. if it is verified that

$$\boldsymbol{\Psi} \tilde{\boldsymbol{\Psi}}^H = \mathbf{I}, \tag{7.54}$$

being $\mathbf{I}$ the identity matrix, it follows the important result that analysis and synthesis operations in plenacoustic transform are linear operations involving the elements of the frame (ψ, a, M) and of its dual $(\tilde{\psi}, a, M)$. More formally, we can write [Moreno-Picot et al., 2010, Eqs. 9-10]

$$\tilde{\mathbf{p}} = \tilde{\boldsymbol{\Psi}}^H \mathbf{p} \tag{7.55}$$

$$\mathbf{p} = \boldsymbol{\Psi} \tilde{\mathbf{p}}. \tag{7.56}$$

According to the choice of the parameters M and N denoting, respectively, the number of considered directions in the ray space and the number of sub-array partitions, we can discriminate between the following two cases.

- If $MN = L$, the matrices $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$ are square and $\tilde{\boldsymbol{\Psi}}^H$ is the inverse of $\boldsymbol{\Psi}$; the condition $MN = L$ is usually referred to as *critical sampling*.
- If $MN > L$, there exist an infinite number of matrices satisfying (7.54); this case is usually referred to as *oversampled* case. One of the possible choices for $\tilde{\boldsymbol{\Psi}}$ is the pseudo-inverse of $\boldsymbol{\Psi}$, i.e.

$$\tilde{\boldsymbol{\Psi}}_0 = (\boldsymbol{\Psi} \boldsymbol{\Psi}^H)^{-1} \boldsymbol{\Psi} \tag{7.57}$$

The following definition for the plenacoustic transform summarizes the derivation in this paragraph.

Definition 10 (Plenacoustic transform)**.** The *plenacoustic transform* of an acoustic field $\mathbf{p}$ sampled by a microphone array is

$$\tilde{\mathbf{p}} = \tilde{\boldsymbol{\Psi}}^H \mathbf{p}. \tag{7.58}$$

The *inverse plenacoustic transform transform* of $\tilde{\mathbf{p}}$ is

$$\mathbf{p} = \boldsymbol{\Psi} \tilde{\mathbf{p}}. \tag{7.59}$$

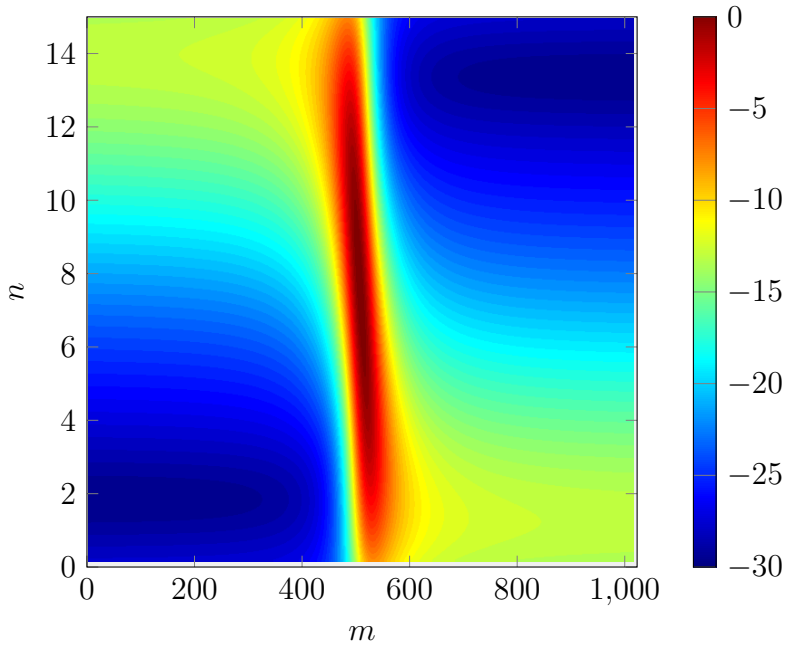


Figure 7.13: Magnitude of the plenacoustic transform $|\tilde{\mathbf{P}}(\omega)|$.

7.5.2 Interpretation

In the light of the discussion in Sec. 4.4, the plenacoustic coefficients obtained through (7.58) have a straightforward physical interpretation. Indeed, the coefficients $\tilde{\mathbf{P}}_{m,n}(\omega)$ represents the complex amplitude of a beam wave object originating at $z = ma$ and whose beam axis is directed as $\theta_n = \arccos(nb)$.

In order to clarify this point, Fig. 7.13 shows the absolute value squared of the coefficients $\tilde{\mathbf{P}}(\omega)$ resulting from the analysis of an acoustic field generated by a point source in $[0.5\,\mathrm{m}, 0, 0]^T$. The array is composed by $L = 16$ microphones, with an inter-element distance $d = 5\,\mathrm{cm}$. The speed of sound is $c = 340\,\mathrm{m\,s^{-1}}$. In Fig. 7.13, for the purpose of showing the analogy with plenacoustic images, the coefficients are mapped on the ray space through the transformation (7.40). The Gaussian window used for analysis is set to

$$\psi[l] = e^{-\frac{l^2}{\sigma_1^2}}, \tag{7.60}$$

with $\sigma_1 = 5$. The Gabor parameters are set to $a = 1$ and $M = 1024$, resulting in a number of windows $N = 16$.

It is evident that Gabor representation does not share the same goal as plenacoustic representation. While the latter is aimed at providing in informative image of the acoustic field, that can be analyzed in order to retrieve information about the acoustic field, the first is aimed at representing the information carried by the acoustic field in a format that is more suitable for processing. For instance, filtering operations can be designed in the plenacoustic domain in order to enhance or attenuate specific local directional components of the acoustic field.

The task of processing acoustic field in the plenacoustic domain requires the plenacoustic transform to be invertible. For this purpose, it is important to notice that Gabor frames (on top of which the plenacoustic transform is built) exhibit perfect reconstruction, in the sense that the subsequent application of the analysis operation (7.58) and the synthesis operation (7.59) does not introduce loss of information.

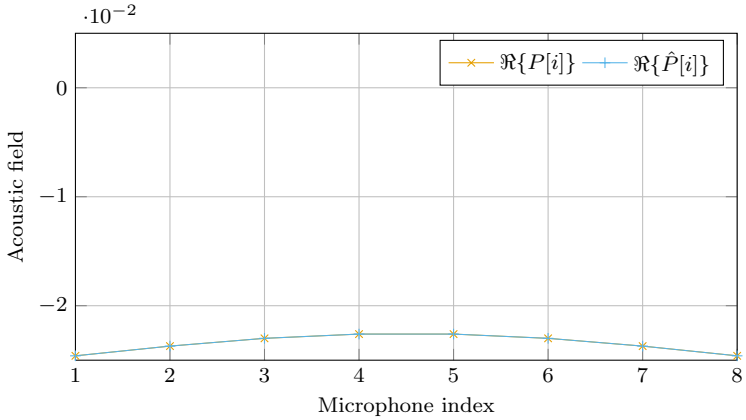


Figure 7.14: Acoustic field reconstruction at microphone positions using plenacoustic analysis and synthesis.

7.5.3 Invertibility of the Plenacoustic Transform

In this paragraph we present simulations to prove the invertibility of the plenacoustic transform defined in Def. 10. The property of invertibility as also referred to as *perfect reconstruction* since it implies that the subsequent application of (7.58) and (7.59) to the acoustic field $\mathbf{p}$ yields exactly $\mathbf{p}$ itself.

In order to provide a simulative validation of the invertibility property, consider the following scenario. The acoustic field has been sampled by a linear array of $I = 8$ microphones displaced on the z axis as in Fig. 7.10. The length of the array is 70 cm and the distance between microphones is $d = 10\,\text{cm}$. The parameters of the discrete Gabor frame are set to $a = 1$ and $M = 30$. The temporal frequency is set to $\omega = 2\pi 500\,\text{Hz}$ and the source is placed in $\mathbf{r} = [3\,\text{m}, 0, 0]^T$. Figure 7.14 shows the real part of the original acoustic field $\mathbf{p}$ and the real part of the acoustic field $\hat{\mathbf{p}}$ reconstructed through plenacoustic analysis and synthesis. Figure 7.14 qualitatively shows that plenacoustic analysis and synthesis exhibits perfect reconstruction. To quantitatively support this claim, consider the normalized mean squared error

$$\bar{\varepsilon} = \frac{1}{I} = \sum_{i=0}^{I-1} \frac{|P[i] - \hat{P}[i]|^2}{|P[i]|^2}. \tag{7.61}$$

In the simulative setup used to compute the data shown in Fig. 7.14 it results $\bar{\varepsilon} = 4.13 \cdot 10^{-32}$ and $10 \log_{10}(\varepsilon) = -313\,\text{dB}$. This means that even in a challenging scenario with few coefficients at hand the proposed transform is invertible without loss of information.

7.6 Main Results

- Sec. 7.2 presents a technique for localizing acoustic source in the vicinity of the microphone array based on the plenacoustic representation. A novel method is presented to overcome resolution issues. Furthermore, the plenacoustic representation is used to extract in an intuitive fashion the radiance pattern of acoustic sources.
- Sec. 7.3 employs the source localization technique previously introduced in a more complex acoustic field analysis methodology aimed at extracting the signal emitted by a desired acoustic source while attenuating interferer and noise.

Chapter 8

Conclusions and Future Works

8.1 Conclusions

In this thesis we have proposed a unified framework for dealing with a wide range of acoustic signal analysis, synthesis and processing tasks. We have provided a comprehensive view of recent advances in signal processing, acoustic theory and representation theory and on top of these concepts we have provided advancements with respect to the state of the art.

The first contribution of this thesis lies in the attempt of promoting among the signal processing community a novel representation for acoustic field based on beams, defined as spatially localized wave objects. We have introduced such wave objects from a physical perspective and we have shown how to represent an arbitrary acoustic field as a summation of them.

In Chap. 5 we have proposed a technique for the accurate estimation of acoustic reflections in a based on impulse responses captured by a spherical microphone array. In this context, we have introduced a twofold modification of state-of-the-art schemes: at first, we have employed an explicit model for the scattering due to the rigid sphere that hosts the microphone array; then, we have shown that it is possible to adopt state of the art approaches from the literature of spectral analysis in order to detect the directional distribution of acoustic energy impinging on the array. We have validated the proposed approach in an acoustically controlled environment and in an auditorium.

In Chap. 6 we have proposed an analytic approach to the problem of acoustic field synthesis with loudspeaker arrays, based on the plane-wave representation of the desired acoustic field. We have derived analytical expressions of the loudspeaker weights for a circular loudspeaker array, providing insight on the reproduction error. Furthermore, we have investigated the relationship between the number of loudspeakers, the size of the reproduction region, the frequency range of operation, and the desired accuracy. We have then shown that the geometrical reasoning enabled by the analogy between plane waves and acoustic rays allows us to easily extend the proposed approach to arbitrarily shaped rendering systems. We ha provided simulations aimed at assessing the effectiveness of the proposed rendering approach in reproducing a desired acoustic field for different rendering geometries (circular, elliptical, and linear); we have provided comparisons with state-of-the-art rendering techniques.

Moreover, in Chap. 6 we have presented a methodology for the design of spatial filters aimed at reproducing plane waves through loudspeakers under uncertainties in the propagation characteristics of the acoustic drivers. Experimental results and simulations showed that the new technique that we have introduced is more robust than the traditional

Tikhonov regularization. Finally, we have applied such robust technique to the rendering approach described in the first sections of Chap. 6 and we have shown a significant improvement of the performance.

In Chap. 7 we have proposed several innovations with respect to the state of the art. First, we have proposed a technique for the deconvolution of plenacoustic images, aimed at improving the resolution capabilities. The deconvolution is based on the fitting of the covariance matrix of array data. In order to reduce the computational cost, we propose a modification of the algorithm, which takes advantage of the information contained in the plenacoustic image. By applying the deconvolution operation, novel applications of plenacoustic imaging are possible, such as the estimation of the radiance pattern of acoustic sources.

Moreover, we have proposed a technique for the extraction of sound sources from convolutive mixtures captured by a microphone array. The plenacoustic image is exploited in order to gather information about the spatial distribution of desired and unwanted acoustic sources. A set of spatial filters is applied to the microphone array in order to enhance or attenuate the desired or undesired sources, while minimizing diffuse field and microphone self noise. We have shown that the adoption of the plenacoustic framework simplifies the whole process. The effectiveness of the proposed solution is validated through simulations and numerical evaluation.

The most innovative contribution of this thesis is the definition of the plenacoustic transform in the setting of discrete Gabor frames. We have derived closed form linear operators to compute both analysis and synthesis operations in the plenacoustic transform. Moreover, we have proved that the transform is invertible without loss of information. We have also provided a physical interpretation for the plenacoustic coefficients in transformed domain, since they encode the complex amplitude of beam wave objects whose origins and axes are determined by the plenacoustic system itself. con

8.2 Future Work

Future works are mainly concentrated in the area of local analysis and synthesis of acoustic field. First, we envision to define filtering operations in the plenacoustic domain; such filtering operations will exploit the highly-structured nature of the plenacoustic domain in order to facilitate the filter design process.

As a second activity, we plan to investigate how the choice of the window function and of its parameters can be used to control both analysis and synthesis processes, trying to understand how they affect the resolution, the depth of field, the invertibility, etc. We envision that multi-resolution strategies can benefit both for analysis and synthesis purposes.

Then, following the route traced with the application of the deconvolution operation to the process of plenacoustic imaging, we aim at improving the acuity of the plenacoustic transform, even exploiting data coming from several microphone arrays.

Finally, we plan activities to devise fast algorithms for the computation of the plenacoustic transform, suitable for low-cost computational platforms and ready to be accommodated in distributed computing paradigms.

Bibliography

T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1655–1666, Aug. 2010.

T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2002.

T. D. Abhayapala, R. A. Kennedy, and R. C. Williamson, "Nearfield broadband array design using a radially invariant modal expansion," *J. Acoust. Soc. Am.*, vol. 107, no. 1, pp. 392–403, Jan. 2000.

M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, 10th ed. Washington DC, USA: National Bureau of Standards, 1972.

W. B. Adams, J. P. Kuhn, and W. P. Whyland, "Correlator compensation requirements for passive time delay estimation with moving source or receivers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 2, pp. 158–168, Apr. 1980.

J. Ahonen, V. Sivonen, and V. Pulkki, "Parametric spatial sound processing applied to bilateral hearing aids," in *Proc. 45th AES Int. Conf. on Applications of Time-Frequency Processing of Audio*, 2008.

J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Berlin, DE: Springer-Verlag, 2012.

J. Ahrens and S. Spors, "Implementation of directional sources in wave field synthesis," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2007.

——, "Rendering of virtual sound sources with arbitrary directivity in higher order ambisonics," in *Proc. AES 123rd Conv.*, 2007.

——, "Analytical driving functions for higher order ambisonics," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2008.

——, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, Nov./Dic. 2008.

——, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2038–2050, Nov. 2010.

——, "A modal analysis of spatial discretization of spherical loudspeaker distributions used for sound field synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2564–2574, Nov. 2012.

T. Ajdler, "The plenacoustic function and its applications," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2006.

T. Ajdler and M. Vetterli, "Acoustic based rendering by interpolation of the plenacoustic function," in *Proc. SPIE/IS&T Visual Communications and Image Processing Conf.*, 2003.

——, "The plenacoustic function, sampling and reconstruction," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2003.

T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Interpolation of head related transfer functions considering acoustics," in *Proc. AES 118th Conv.*, 2005.

T. Ajdler, L. Sbaiz, A. Ridolfi, and M. Vetterli, "On a stochastic version of the plenacoustic function," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2006.

T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

A. Alexandridis, A. Griffin, and A. Mouchtaris, "Capturing and reproducing spatial audio based on a circular microphone array," *J. Electrical and Computer Engineering*, vol. 2013, pp. 1–16, 2013.

J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-25, no. 3, pp. 235–238, 1977.

R. A. Altes, "Target position estimation in radar and sonar, and generalized ambiguity analysis for maximum likelihood parameter estimation," *Proceedings of the IEEE*, vol. 67, no. 6, pp. 920–930, June 1979.

P. Annibale, "Accuracy of time-difference-of-arrival based source localization algorithms under temperature variations," in *Proc. Int. Symposium on Communications, Control and Signal Processing*, 2010.

P. Annibale and R. Rabenstein, "Acoustic source localization and speed estimation based on time-differences-of-arrival under temperature variations," in *Proc. European Signal Processing Conference*, 2010.

——, "Speed of sound in air temperature estimation using TDOA-based localization framework," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2012.

P. Annibale, R. Rabenstein, S. Spors, and P. Steffen, "A short review of signals and systems for spatial audio," in *Proc. European Signal Processing Conf.*, 2009.

F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.

F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2009.

F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.

G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed. Burlington, MA, USA: Elsevier Academic Press, 2005.

M. Askari, M. Karimi, and Z. Atbaee, "Robust beamforming in circular arrays using phase-mode transformation," *IET Signal Processing*, vol. 7, no. 8, pp. 693–703, 2013.

J. Atkins, "Robust beamforming and steering of arbitrary beam patterns using spherical arrays," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2011.

M. R. Azimi-Sadjadi, A. Pezeshki, L. L. Scharf, and M. Hohil, "Wideband DOA estimation algorithms for multiple target detection and tracking using unattended acoustic sensors," in *Proc. SPIE 5417, Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI*, 2004.

R. Balan, P. G. Casazza, and D. Edidin, "On signal reconstruction without phase," *J. App. Comput. Harmonic Anal.*, vol. 20, no. 3, pp. 345–356, Mar. 2006.

J. J. Benedetto, A. M. Powell, and O. Yilmaz, "Sigma-delta quantization and finite frames," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1990–2005, May 2006.

J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Berlin, DE: Springer-Verlag, 2005.

J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.

J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, DE: Springer-Verlag, 2008.

J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, ser. Springer Briefs in Electrical and Computer Engineering. Heidelberg, DE: Springer, 2012.

A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

R. Bernardini and R. Rinaldo, "Efficient reconstruction from frame-based multiple descriptions," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3282–3296, Aug. 2005.

——, "Bounds on error amplification in oversampled filter banks for robust transmission," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1399–1411, Apr. 2006.

——, "Oversampled filter banks from extended perfect reconstruction filter banks," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp. 2625–2635, July 2006.

P. Bestagini, M. Compagnoni, F. Antonacci, A. Sarti, and S. Tubaro, "TDOA-based acoustic source localization in the space-range reference frame," *Multidimensional Systems and Signal Processing*, vol. 25, no. 3, pp. 337–359, Mar. 2013.

L. Bianchi, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Deconvolution of plenacoustic images," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2013.

L. Bianchi, R. Magalotti, F. Antonacci, A. Sarti, and S. Tubaro, "Robust beamforming under uncertainties in the loudspeaker directivity pattern," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2014.

L. Bianchi, F. D'Amelio, F. Antonacci, A. Sarti, and S. Tubaro, "A plenacoustic approach to acoustic signal extraction," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2015.

L. Bianchi, M. Verdi, F. Antonacci, A. Sarti, and S. Tubaro, "High resolution imaging of acoustic reflections with spherical microphone array," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2015.

L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "Model-based acoustic rendering based on plane wave decomposition," *Applied Acoustics*, vol. 104, pp. 127–134, Mar. 2016.

S. Boyd and L. Vandenberghe, *Convex Optimization*, 7th ed. Cambridge, MA, USA: Cambridge University Press, 2009.

R. N. Bracewell, *The Fourier Transform and its Applications*, 3rd ed. Singapore: McGraw-Hill, 2000.

P. Brémaud, *Mathematical Principles of Signal Processing.* New York, NY, USA: Springer Science+Business Media, 2002.

T. F. Brooks and J. W. M. Humphreys, "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays," in *Proc. AIAA/CEAS Aeroacoustics Conf.*, Manchester, UK, May 10-12 2004.

H. Buchner, S. Spors, W. Kellermann, and R. Rabenstein, "Full-duplex communication system with loudspeaker arrays and microphone arrays," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2002.

P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2011.

A. Canclini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Bucharest, RO, Aug. 27-31 2012.

A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 439–443, Feb. 2013.

A. Canclini, D. Marković, L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "A geometrical approach to room compensation for sound field rendering," in *Proc. 22nd European Signal Processing Conference*, 2014.

——, "A robust geometric approach to room compensation for sound field rendering," *IEICE Trans. Fundamentals*, vol. E97, pp. 1884–1892, 2014.

A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1563–1575, Oct. 2015.

J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

A. Carini, S. Cecchi, F. Piazza, I. Omiciuolo, and G. L. Sicuranza, "Multiple position room response equalization in frequency domain," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 122–135, Jan. 2012.

G. C. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 3, pp. 463–470, June 1981.

S. Cecchi, A. Primavera, F. Piazza, and A. Carini, "Mixed time-frequency approach for multipoint room response equalization," in *Proc. 45th AES Int. Conf.*, 2012.

S. Cecchi, L. Romoli, F. Piazza, and A. Carini, "A multichannel and multiple position adaptive room response equalizer in warped domain," in *Proc. Int. Symposium on Image and Signal Processing and Analysis*, 2013.

S. Cecchi, M. Virgulti, F. Bettarelli, S. Doria, and F. Piazza, "Development of multipoint mixed-phase equalization system for multiple environments," in *Proc. 134th AES Conv.*, 2013.

S. Cecchi, A. Primavera, F. Piazza, F. Bettarelli, and J. Li, "Advanced audio spatializer combined with a multipoint equalization system," in *Proc. IEEE World Congress on Computational Intelligence*, 2014.

S. Cecchi, L. Romoli, A. Carini, and F. Piazza, "A multichannel and multiple position adaptive room response equalizer in warped domain: Real-time implementation and performance evaluation," *Applied Acoustics*, vol. 82, pp. 28–37, Aug. 2014.

S. Cecchi, L. Romoli, F. Piazza, B. Bank, and A. Carini, "A novel approach for prototype extraction in a multipoint equalization procedure," in *Proc. 136th AES Conv.*, 2014.

Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

G. Cincotti, F. Gori, M. Santarsiero, F. Frezza, F. Furnò, and G. Schettini, "Plane wave expansion of cylindrical functions," *Optics Communications*, vol. 95, pp. 192–198, 1993.

D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory.* Berlin Heidelberg, DE: Springer-Verlag, 1992.

M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources through the fitting of propagation cones," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 1964–1975, 2012.

M. Compagnoni, R. Notari, F. Antonacci, and A. Sarti, "A comprehensive analysis of the geometry of TDOA maps in localization problems," *Inverse Problems*, vol. 30, pp. 1–49, 2014.

C. E. Cook and M. Bernfeld, *Radar Signals.* New York, NY, USA: Academic Press, 1967.

R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, "Measurement of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.*, vol. 27, pp. 1072–1077, 1955.

E. Corteel, "Synthesis of directional sources using wave field synthesis, possibilities, and limitations," *J. Advances in Signal Processing*, vol. 2007, pp. 1–18, 2007.

M. Costa, A. Richter, and V. Koivunen, "Unified array manifold decomposition based on spherical harmonics and 2-D fourier basis," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4634–4645, Sept. 2010.

Z. Cvetković, "Resilience properties of redundant expansions under additive noise and quantization," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 644–656, Mar. 2003.

I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format," in *Proc. 23rd AES Int. Conf.*, 2003.

I. Daubechies and R. DeVore, "Reconstructing a bandlimited function from very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order," *Ann. Math.*, vol. 158, no. 2, pp. 679–710, Feb. 2003.

A. J. Devaney and G. C. Sherman, "Plane-wave representation for scalar wave fields," *SIAM Review*, vol. 15, no. 4, pp. 765–776, Oct. 1973.

A. J. Devaney and E. Wolf, "Multipole expansions and plane wave representations of the electromagnetic field," *J. Mathematical Physics*, vol. 15, no. 2, pp. 234–244, Feb. 1974.

S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.

I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *PNAS*, vol. 110, no. 30, pp. 12 186–12 191, 2013.

P. L. Dragotti, V. Velisavljevic, M. Vetterli, and B. Beferull-Lozano, "Discrete directional wavelet bases and frames for image compression and denoising," in *Proc. SPIE Conf. Wavelet Applications Signal Image Processing*, 2003, pp. 1287–1295.

R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2005.

N. Epain and C. T. Jin, "Super-resolution sound field imaging with sub-space preprocessing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2013.

A. A. Ergin, B. Shanker, and E. Michielssen, "The plane-wave time-domain algorithm for the fast analysis of transient wave phenomena," *IEEE Antennas and Propagation Magazine*, vol. 41, no. 4, pp. 39–52, Aug. 1999.

Sound Quality Assessment Material recording for subjective tests, European Broadcasting Union Std., 2008.

F. Fahy, *Sound Intensity*, 2nd ed. London, UK: E & FN Spon, 1995.

C. Faller, "Parametric coding of spatial audio," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2004.

A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Conv.*, Feb. 2000.

A. Farina, A. Capra, L. Chiesi, and L. Scopece, "A spherical microphone array for synthesizing virtual directive microphones in live bradcasting and in post production," in *Proc. AES 140th Int. Conf.*, 2010.

A. Farina, A. Amendola, A. Capra, and C. Varani, "Spatial analysis of room impulse responses captured with a 32-capsules microphone array," in *Proc. AES 130th Conv.*, 2011.

F. Fazi and P. Nelson, "The ill-conditioning problem in sound field reconstruction," in *Proc. 123rd Conv. Audio Eng. Soc.*, New York, NY, USA, 2007.

F. M. Fazi and P. A. Nelson, "Nonuniqueness of the solution of the sound field reproduction problem with boundary pressure control," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 1–14, Jan./Feb. 2012.

H. G. Feichtinger, W. Kozek, and F. Luef, "Gabor analysis over finite abelian groups," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 2, pp. 230–248, Feb. 2009.

P. Filippi, A. Bergassoli, D. Habault, and J. Lefebvre, *Acoustics. Basic Physics, Theory, and Methods*. London, UK: Academic Press, 1998.

S. Fiori and F. Piazza, "Neural MCA for robust beamforming," in *Proc. IEEE Int. Symposium on Circuits and Systems*, 2000.

E. Fisher and B. Rafaely, "Near-field spherical microphone array processing with radial filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 256–265, Feb. 2011.

S. Foster, "Impulse response measurement using golay codes," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Tokyo, JP, Apr. 8–11, 1986, pp. 929–932.

O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

D. Gabor, "Theory of communication. part 1: The analysis of information," *J. Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, Nov. 1946.

P. A. Gauthier, C. Camier, Y. Pasco, A. Berry, E. Chambatte, R. Lapointe, and M. A. Delalay, "Beamforming regularization matrix and inverse problems applied to sound field measurement and extrapolation using microphone array," *J. Sound and Vibration*, vol. 330, pp. 5852–5877, 2011.

M. Gillette and H. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, Jan. 2008.

P. F. Goldsmith, *Quasioptical Systems*, ser. Series on microwave technology and techniques. New York, NY, USA: IEEE Press / Chapman & Hall Publishers, 1998.

J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. New York, NY, USA: McGraw-Hill, 1996.

M. Goodwin and J. M. Jot, "Spatial audio scene coding," in *Proc. 125th AES Conv.*, Oct. 2008.

B. N. Gover, "Directional measurement of airborne sound transmission paths using a spherical microphone array," *J. Audio Eng. Soc.*, vol. 53, no. 9, pp. 787–795, Sept. 2005.

B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound field," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 1980–1991, Nov. 2002.

——, "Measurement of directional properties of reverberant sound fields in rooms using a spherical microphone array," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2138–2148, Apr. 2004.

V. K. Goyak, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in R^N: Analysis, synthesis, and algorithms," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 16–31, Jan. 1998.

I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Burlington, MA, USA: Academic Press, 2007.

M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag, 2008, pp. 95–110.

M. Guillaume and Y. Grenier, "Sound field analysis based on analytical beamforming," *J. Advances in Signal Processing*, vol. 2007, pp. 1–15, 2007.

B. Günel, H. Hacihabiboglu, and A. M. Kondoz, "Plane wave decomposition with regularization using a single rotating microphone," in *Proc. AES 30th Int. Conf.*, Saariselka, FI, March 15-17 2007.

A. Gupta and T. D. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1149–1159, July 2011.

J. Hannemann and K. D. Donohue, "Virtual sound source rendering using a multipole-expansion and method-of-moments approach," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 473–481, June 2008.

T. B. Hansen, "Exact plane-wave expansion with directional spectrum: Application to transmittting and receiving antennas," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 8, pp. 4187–4198, Aug. 2014.

T. B. Hansen and A. D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields.* New York, NY, USA: IEEE Press, 1999.

T. Heilpern and E. Heyman, "Born approximate inverse scattering using beam basis functions," in *Proc. Int. Symposium on Electromagnetic Theory*, 2013.

——, "MUSIC imaging using phase-space gaussian-beams processing," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1270–1281, Mar. 2014.

W. Herbordt, H. Buchner, W. Kellermann, R. Rabenstein, S. Spors, and H. Teutsch, "Full-duplex multichannel communication: real-time implementations in a general framework," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, 2003.

J. Herre, C. Falch, D. Mahne, G. D. Galdo, M. Kallinger, and O. Thiergart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," *J. Audio Eng. Soc.*, vol. 59, no. 12, pp. 924–935, Dec. 2011.

Y. Hioka and T. Betlehem, "Under-determined source separation based on power spectral density eestimate using cylindrical model beamforming," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2013.

F. M. Hoffmann and F. M. Fazi, "Theoretical study of acoustic circular arrays with tangential pressure gradient sensors," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1762–1774, Nov. 2015.

N. Huleihel and B. Rafaely, "Spherical array processing for acoustic analysis using room impulse responses and time-domain smoothing," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3995–4007, Jun. 2013.

C. T. Jin, N. Epain, and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 193–204, Jan. 2014.

D. H. Johnson and D. E. Dudgeon, *Array Signal Processing Concepts and Techniques.* Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

R. A. Kennedy and P. Sadeghi, *Hilbert Space Methods in Signal Processing.* Cambridge, UK: Cambridge University Press, 2013.

R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Transactions on Signal Processing*, vol. 46, no. 8, pp. 2147–2156, Aug. 1998.

R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2542–2556, June 2007.

Z. Khalid, R. A. Kennedy, and J. D. McEwen, "An optimal-dimensionality sampling scheme on the sphere with fast spherical harmonic transforms," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4597–4610, Sept. 2014.

D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 261–270, July 2012.

L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. New York, NY, USA: John Wiley & Sons, 2000.

O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, Nov. 1993.

O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1584–1593, Sept. 1996.

H. Kogelnik and T. Li, "Laser beams and resonators," *Proceedings of the IEEE*, vol. 54, no. 10, pp. 1312–1330, Oct. 1966.

M. Kolundzija, C. Faller, and M. Vetterli, "Designing practical filters for sound field reconstruction," in *Proc. AES 127th Conv.*, New York, NY, USA, Oct. 9-12 2009.

——, "Sound field reconstruction: An improved approach for wave field synthesis," in *Proc. AES 126th Conv.*, Munich, DE, May 7-10 2009.

——, "Design of a compact cylindrical loudspeaker array for spatial sound reproduction," in *Proc. AES 130th Conv.*, London, UK, May 13-16 2011.

A. Koretz and B. Rafaely, "Dolph-chebyshev beampattern design for spherical arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2417–2420, June 2009.

J. Kovacevic and A. Chebira, "Life beyond bases: The advent of frames (part I)," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 86–104, July 2007.

K. Kowalczyk, O. Thiergart, M. Taseska, G. D. Galdo, V. Pulkki, and E. A. P. Habets, "Parametric spatial sound processing," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 31–42, Mar. 2015.

M. Kubovy and D. V. Valkenburg, "Auditory and visual objects," *Cognition*, vol. 80, pp. 97–126, 2001.

K. Kumatani, L. Lu, J. McDonough, A. Ghoshal, and D. Klakow, "Maximum negentropy beamforming with superdirectivity," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aalborg, DK, Aug. 23-27 2010.

A. Kuntz, "Wave field analysis using virtual circular microphone arrays," Ph.D. dissertation, Friedrich-Alexander-Universität Erlanger-Nürnberg, 2008.

A. Kuntz and R. Rabenstein, "Limitations in the extrapolation of wave fields from circular measurements," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Poznan, PL, Sept. 3.7 2007.

H. Kuttruff, *Room Acoustics*. Oxon, UK: Spon Press, 2009.

C. Lai, S. Nordholm, and Y. Leung, "Design of robust steerable broadband beamformers incorporating microphone gain and phase error characteristics," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2011.

C. C. Lai, S. Nordholm, and Y. H. Leung, "Design of steerable spherical broadband beamformers with flexible sensor configurations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 427–438, 2013.

M. V. Laitinen and V. Pulkki, "Binaural reproduction for directional audio coding," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, Oct. 2009, pp. 337–340.

E. Lalor, "Conditions of the validity of the angular spectrum of plane waves," *J. Optical Soc. Am.*, vol. 58, no. 9, pp. 1235–1237, 1968.

J. M. Lee, Y. W. Choi, and Y. H. Kim, "Reproduction of a higher-order circular harmonic field using a linear array of microphones," *J. Acoust. Soc. Am. Express Letters*, vol. 137, no. 3, pp. EL227–EL233, Mar. 2015.

M. Legg and S. Bradley, "A combined microphone and camera calibration technique with application to acoustic imaging," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4028–4039, 2013.

D. Levin, E. A. P. Habets, and S. Gannot, "Robust beamforming using sensors with nonidentical directivity pattern," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2013.

J. Li and H. Chen, "Least squares frequency invariant beamforming robust against microphone mismatches," in *Proc. Int. Conf. on Information Science and Technology*, 2011.

Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 702–714, Feb. 2007.

G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.

J. Lorente, G. Pinero, A. M. Vidal, J. A. Belloch, and A. Gonzalez, "Parallel implementations of beamforming design and filtering for microphone array applications," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Barcelona, ES, Aug. 29 - Sept. 2 2011.

D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY, USA: John Wiley & Sons, 1969.

D. Lugara, C. Letrou, A. Shlivinski, E. Heyman, and A. Boag, "Frame-based gaussian beam summation method: Theory and applications," *Radio Science*, vol. 38, no. 2, pp. 1–15, 2003.

E. Mabande, A. Schad, and W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2009.

——, "A time-domain implementation of data-independent robust broadband beamformers with low filter order," in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, 2011.

J. J. Maciel and L. B. Felsen, "Discretized gabor-based beam algorithm for time-harmonic radiation from two-dimensional truncated planar aperture distributions - i: Formulation and solution," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 12, pp. 1751–1759, Dec. 2002.

R. H. MacPhie and K. L. Wu, "A plane wave expansion of spherical wave functions for modal analysis of guided wave structures and scatterers," *IEEE Transactions on Antennas and Propagation*, vol. 51, pp. 2801–2805, 2003.

L. Mandel and E. Wolf, *Optical coherence and quantum optics*. Cambridge, UK: Cambridge University Press, 1995.

D. Marković, G. Sandrini, F. Antonacci, A. Sarti, and S. Tubaro, "Plenacoustic imaging in the ray space," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.

D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Estimation of room dimensions from a single impulse response," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2013.

——, "Soundfield imaging in the ray space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2493–2505, Dec. 2013.

D. Marković, K. Kowalczyk, F. Antonacci, C. Hoffmann, A. Sarti, and W. Kellermann, "Estimation of acoustic reflection coefficients through pseudospectrum matching," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 125–137, Jan. 2014.

D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Multiview soundfield imaging in the projective ray space," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 6, pp. 1054–1067, June 2015.

S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple inferfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

F. Martellotta, "On the use of microphone arrays to visualize spatial sound field information," *Applied Acoustics*, vol. 74, pp. 987–1000, 2013.

J. D. McEwan, M. P. Hobson, D. J. Mortlock, and A. N. Lasenby, "Fast directional continuous spherical wavelet transform algorithms," *IEEE Transactions on Signal Processing*, vol. 55, pp. 520–529, 2007.

T. Melamed, "Phase-space beam summation: a local spectrum analysis of time-dependent radiation," *J. Electromagnetic Waves and Applications*, vol. 11, no. 6, pp. 739–773, 1997.

J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects," *J. Acoust. Soc. Am.*, vol. 109, no. 1, pp. 185–193, Jan. 2001.

J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2002.

Em32 Eigenmike® microphone array release notes., mh acoustics', Apr. 2013.

N. Mitlanoudis and M. E. Davies, "Using beamforming in the audio source separation problem," in *Proc. Int. Symposium on Signal Process. and its Applications*, 2003.

M. J. Mohlenkamp, "A fast transform for spherical harmonics," *J. Fourier Anal. Appl.*, vol. 5, no. 2/3, pp. 159–184, 1999.

A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous helmholtz solutions," *Z. Angew. Math. Phys.*, vol. 62, pp. 809–837, 2011.

S. Moreno-Picot, M. Arevalillo-Herráez, and W. Díaz-Villanueva, "A linear cost algorithm to compute the discrete gabor transform," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2667–2674, May 2010.

P. M. Morse and H. Feshbach, *Methods of Theoretical Physics.* New York, NY, USA: McGraw-Hill, 1953, vol. I.

P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, with errata page, first Princeton University Press ed. Princeton, NJ, USA: Princeton University Press, 1986.

S. Müller and P. Massarani, "Transfer-function measurements with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, 2001.

M. Nieto-Vesperinas, "Incoming and outgoing components of source-free wavefields," *Optics Communications*, vol. 67, no. 6, pp. 391–395, 1988.

A. O'Donovan, R. Duraiswami, and D. Zotkin, "Imaging concert hall acoustics using visual and audio cameras," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2008.

F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson, and P. Otto, "Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions," in *Proc. AES 52nd Int. Conf.*, 2013.

F. W. J. Olver, Ed., *NIST Handbook of Mathematical Functions.* New York, NY, USA: National Institute of Standards and Technology, 2010.

M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 1994.

M. Omura, M. Yada, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Compensating of room acoustic transfer functions affected by change of room temperature," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 1999.

A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1997, international edition.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998, solutions Manual attached.

M. Park and B. Rafaely, "Sound-field analysis by plane wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3094–3103, Nov. 2005.

L. C. Parra, "Steerable frequency-invariant beamforming for arbitrary arrays," *J. Acoust. Soc. Am.*, vol. 119, no. 6, pp. 3839–3847, June 2006.

A. Parthy, C. Jin, and A. van Schaik, "Acoustic holography with a concentric rigid and open spherical microphone array," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2009.

Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2532–2540, 2013.

E. Perrey-Debain, "Plane wave decomposition in the unit disc: Convergence estimates and computational aspects," *J. Computational and Applied Mathematics*, vol. 193, no. 1, pp. 140–156, Aug. 2006.

F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4608–4622, Sept. 2010.

F. Pinto, M. Kolundzija, and M. Vetterli, "Digital acoustics: processing wave fields in space and time using DSP tools," *APSIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1–21, 2014.

F. P. C. Pinto, "Signal processing in space and time - a multidimensional fourier approach," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, 2010.

M. Pirro, S. Squartini, and F. Piazza, "A fixed beamforming based approach for stereophonic audio-conference systems," in *Proc. 131th AES Conv.*, 2011.

M. Pirro, S. Squartini, L. Romoli, and F. Piazza, "Stereophonic hands-free communication system based on microphone array fixed beamforming: Real-time implementation and evaluation," *EURASIP Journal on Audio, Speech and Music Processing*, vol. 2012, no. 26, pp. 1–21, 2012.

M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

——, "Cylindrical harmonic expansion of the sound field due to a rotating source," in *Proc. Int. Congress on Acoustics*, 2010.

M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Higher-order loudspeakers and active compensation for improved 2d sound field reproduction in rooms," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 31–45, Jan./Feb. 2015.

A. Primavera, S. Cecchi, J. Li, and F. Piazza, "Objective and subjective investigation on a novel method for digital reverberator parameters estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 441–452, 2014.

V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, June 2007.

S. Qiu and H. G. Feichtinger, "Discrete gabor structures and optimal representations," *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2258–2268, Oct. 1995.

R. Rabenstein and L. Trautmann, "Multidimensional transfer function models," *IEEE Transactions on Circuits and Systems I-Fundamental Theory and Applications*, vol. 49, no. 6, pp. 852–861, June 2002.

——, "Towards a framework for continuous and discrete multidimensional systems," *Int. J. Appl. Math. Comput. Sci.*, vol. 13, no. 1, pp. 73–85, Jan. 2003.

R. Rabenstein, P. Steffen, and S. Spors, "A tutorial on the representation of two-dimensional wave fields by multidimensional signals," in *Proc. Int. Workshop on Multidimensional Signals*, 2005.

——, "Representation of two-dimensional wave fields by multidimensional signals," *Signal Processing*, vol. 86, pp. 1341–1351, Oct. 2006.

B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.

——, "Analysis and design of spherical micropone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

——, "The spherical-shell microphone array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 740–747, May 2008.

B. Rafaely and D. Khaykin, "Optimal model-based beamforming and independent steering for spherical loudspeaker arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2234–2238, 2011.

B. Rafaely and M. Kleider, "Spherical microphone array beam steering using wigner-d weighting," *IEEE Signal Processing Letters*, vol. 15, pp. 417–420, 2008.

B. Rafaely, I. Balmages, and L. Eger, "High-resolution plane-wave decomposition in an auditorium using a dual radius scanning spherical microphone array." *J. Acoust. Soc. Am.*, vol. 122, no. 5, pp. 2661–2668, Nov. 2007.

G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 711–727, May 2008.

F. Ribeiro, C. Zhang, D. A. Florêncio, and D. E. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1781–1792, Sept. 2010.

F. P. Ribeiro and V. H. Nascimento, "Fast transform for acoustic imaging - part i. theory," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2229–2240, Aug. 2011.

M. Rübsamen and A. B. Gershman, "Robust adaptive beamforming based on multidimensional covariance fitting," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2010.

——, "Robust adaptive beamforming using multidimensional covariance fitting," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 740–753, Feb. 2012.

M. Rübsamen and M. Pesavento, "Steering vector non-identifiability in covariance matrix fitting based beamforming," in *Proc. Sensor Array and Multichannel Signal Processing Workshop*, 2012.

W. Rudin, *Real and Complex Analysis.* New York, NY, USA: McGraw-Hill, 1987.

J. G. Ryan, "Criterion for the minimum source distance at which plane-wave beamforming can be applied," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 595–598, 1998.

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

P. Samarasinghe, T. D. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 647–658, 2014.

T. Schetelig and R. Rabenstein, "Simulation of three-dimensional sound propagation with multidimensional wave digital filters," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 1998.

R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-8, no. 6, pp. 821–835, Nov. 1972.

E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules," *Physical Review*, vol. 28, no. 6, pp. 1049–1070, Dec. 1926.

R. Schultz-Amling, F. Kuech, O. Thiergart, and M. Kallinger, "Acoustical zooming based on a parametric sound field representation," in *Proc. 128th AES Conv.*, 2010.

T. Sekiguchi and Y. Karasawa, "Wideband beamspace adaptive array utilizing fir filters for multibeam forming," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 277–284, Jan. 2000.

G. C. Sherman, "Diffracted wave fields expressible by plane-wave expansions containing only homogeneous waves," *J. Optical Soc. Am.*, vol. 59, pp. 697–711, 1969.

——, "Plane-wave expansions of the optical field," *Optics Communications*, vol. 6, pp. 115–118, 1972.

A. Shlivinski, E. Heyman, A. Boag, and C. Letrou, "A phase-space beam summation formulation for ultrawide-band radiation," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 8, pp. 2042–2056, Aug. 2004.

A. Shlivinski, E. Heyman, and A. Boag, "A phase-space beam summation formulation for ultrawide-band radiation - part ii: A multiband scheme," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 3, pp. 948–957, Mar. 2005.

Y. Shono and T. Inuzuka, "Representation of a diffracted wave field by the band-limited angular spectrum," *J. Optical Soc. Am.*, vol. 68, pp. 1579–1586, 1978.

J. Smith and J. S. Abel, "Closed-form least-squares source location estimation from frange-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 12, pp. 1661–1669, Dec. 1987.

J. P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections.* Chicago, IL, USA: University of Chicago Press, 1993.

S. Spors, "Active listening room compensation for spatial sound reproduction systems," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2005.

S. Spors and R. Rabenstein, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," in *Proc. AES 120th Conv.*, Paris, FR, May 20-23 2006.

S. Spors, A. Kuntz, and R. Rabenstein, "Listening room compensation for wave field synthesis," in *Proc. Int. Conf. on Multimedia and Expo*, 2003.

S. Spors, H. Teutsch, A. Kuntz, and R. Rabenstein, *Audio Signal Processing fro Next-Generation Multimedia.* Boston, USA: Kluwer Academic Publishers, 2004, ch. Sound Field Synthesis, pp. 323–342.

S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am.*, vol. 122, pp. 354–369, 2007.

S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Proc. AES 124th Conv.*, Amsterdam, NE, May 17-20 2008.

S. Spors, H. Wierstorf, and M. Geier, "Comparison of modal versus delay-and-sum beamforming in the context of data-based binaural synthesis," in *Proc. AES 132nd Conv.*, 2012.

S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial sound with loudspeakers and its perception: A review of the current state," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 1920–1938, 2013.

P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, Nov. 2006.

P. Stoica and R. Moses, *Spectral Analysis of Signals.* Upper Saddle River, NJ, USA: Prentice Hall, 2004.

P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 35–47, Jan. 2011.

——, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

G. Strang, *Linear Algebra and its Applications*, 3rd ed. London, UK: Thomson, 1998.

N. Strobel and R. Rabenstein, "Classification of time delay estimates for robust speaker localization," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 1999.

——, "Robust speaker localization using a microphone array," in *Proc. European Signal Processing Conference*, 2000.

N. Strobel, T. Meier, and R. Rabenstein, "Speaker localization using a steered filter-and-sum beamformer," in *Proc. Workshop on Vision, Modeling, and Visualization*, 1999.

N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 22–31, Jan. 2001.

T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications*. Boston, MA, USA: Birkhäuser, 1998, ch. Numerical algorithms for discrete Gabor expansions, pp. 267–294.

I. Tashev, J. Droppo, M. Seltzer, and A. Acero, "Robust design of wideband loudspeaker arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2008.

S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28, Jan. 2013.

H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2005.

——, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2724–2736, May 2006.

H. Teutsch, S. Spors, W. Herbordt, W. Kellermann, and R. Rabenstein, "An integrated real-time system for immersive audio applications," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2003.

J. Thiemann and E. Vincent, "An experimental comparison of source separation and beamforming techniques for microphone array signal enhancement," in *Proc. Int. Workshop on Machine Learning for Signal Processing*, 2013.

O. Thiergart and E. A. P. Habets, "An informed LCMV filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Vancouver, CA, May 2013.

O. Thiergart, G. D. Galdo, M. Taseska, and E. A. P. Habets, "Geometry-based spatial sound acquisition using distributed microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2583–2594, Dec. 2013.

O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed parametric filter based on instantaneous direction-of-arrival estimates," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2182–2196, Dec. 2014.

M. R. P. Thomas, J. Ahrens, and I. Tashev, "A method for converting between cylindrical and spherical harmonic representations of sound fields," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2014.

B. Treeby, B. Cox, and J. Jaros, *k-Wave A MATLAB toolbox for the time domain simulation of acoustic wave fields*, 2012.

H. L. V. Trees, *Optimum Array Processing.* New York, NY, USA: John Wiley & Sons, 2002, part IV of Detection, Estimation and Modulation Theory.

A. Trucco, M. Crocco, and S. Repetto, "A stochastic approach to the synthesis of a robust frequency-invariant filter-and-sum beamformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1407–1415, Aug. 2006.

B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding.* Englewood Cliffs, NJ, USA: Prentice Hall, 1995.

M. Vetterli, J. Kovacevic, and V. K. Goyal, *Foundations of Signal Processing.* Cambridge, UK: Cambridge University Press, 2014.

E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1669, July 2006.

E. Vincent, S. Araki, F. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, V. Gowreesunker, D. Lutter, and N. Q. K. Dong, "The signal separation evaluation campaign (2007-2010): Achievements and remaining challenges," *Signal Processing*, vol. 92, pp. 1928–1936, 2012.

M. Vorländer, *Auralization*, 1st ed. Berlin, DE: Springer-Verlag, 2008.

D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1023–1034, Feb. 1995.

D. B. Ward, Z. Ding, and R. A. Kennedy, "Broadband DOA estimation using frequency invariant beamforming," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1463–1469, May 1998.

H. W. Wei and S. F. Ye, "Comments on "a linear closed-form algorithm fro source localization from time-differences of arrival"," *IEEE Signal Processing Letters*, vol. 15, p. 895, Nov. 2008.

T. Werther, Y. C. Eldar, and N. K. Subbanna, "Dual gabor frames: Theory and computational aspects," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 4147–4158, Nov. 2005.

H. Weyl, "Ausbreitung elektromagnetischer wellen über einen ebenen leiter," *Ann. Phys.*, vol. 60, pp. 481–500, 1919.

E. T. Whittaker, "On the partial differential equations of mathematical physics," *Math. Ann.*, vol. 57, pp. 333–355, 1903.

E. G. Williams, *Fourier Acoustics*. London, UK: Academic Press, 1999.

P. M. Woodward, *Probability and Information Theory with Applications to Radar*. Oxford, UK: Pergamon Press, 1964.

Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 107–116, 2009.

Y. Xu, J. B. Weaver, J. D. M. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 747–758, Dec. 1994.

S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 361–371, Feb. 2011.

T. Yardibi, J. Li, P. Stoica, and L. N. Cattafesta, "Sparsity constrained deconvolution approaches for acoustic source mapping," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 2631–2642, May 2008.

T. Yardibi, J. Li, P. Stoica, N. S. Zawodny, and L. N. Cattafesta, "A covariance fitting approach for correlated acoustic source mapping," *J. Acoust. Soc. Am.*, vol. 127, no. 5, pp. 2920–2931, May 2010.

L. Zamaninezhad, P. Annibale, and R. Rabenstein, "Localization of environmental reflectors from a single measured transfer function," in *Proc. IEEE Int. Symposium on Communications, Control and Signal Processing*, 2014.

M. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Empirical determination of frequency representation in spherical harmonics-based HRTF functional modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 351–360, Feb. 2015.

D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Sound field decomposition using spherical microphone arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2008.

——, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, pp. 2–16, Jan. 2010.

F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," Ph.D. dissertation, University of Music and Performing Arts, Gratz, 2009.