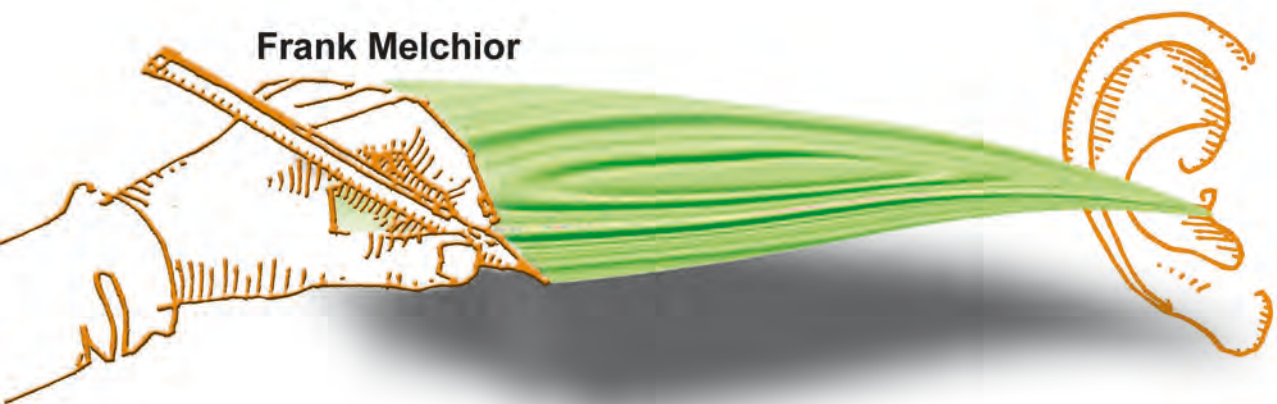


spatial
SOUND design
BASED ON MEASURED ROOM IMPULSE RESPONSES

Frank Melchior



Investigations on spatial sound design
based on
measured room impulse responses

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 24 juni 2011 om 15:00 uur

door

Frank MELCHIOR

Diplom - Ingenieur, Technische Universität Ilmenau,
geboren te Wesel, Duitsland.

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. A. Gisolf

Copromotor: Dr. ir. D. de Vries

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. A. Gisolf,	Technische Universiteit Delft, promotor
Dr. ir. D. de Vries,	Technische Universiteit Delft, copromotor
Prof. dr. dr. ir. K.H. Brandenburg,	Technische Universität Ilmenau
Prof. dr. S. Weinzierl,	Technische Universität Berlin
Prof. dr. ir. P.P. Jonker,	Technische Universiteit Delft
Prof. dr. C.M. Jonker,	Technische Universiteit Delft
Prof. dr. ir. L.A.M. van der Wielen,	Technische Universiteit Delft

ISBN 978-94-91211-68-3

Copyright 2011, by Frank Melchior. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.

This research was supported by the Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany.

Typesetting system: L^AT_EX

To my parents.

Für meine Eltern.

Preface

We shape our tools
and thereafter they shape us.

John M. Culkin

A strong connection exists between the developments of tools and technology and the progress in creative-artistic disciplines. This holds especially for music and sonic art. This interaction manifests itself from the early investigations of Kircher (1684) starting in the 16th century. Kircher proposes the integration of acoustic effects (e.g., acoustic echo) in the composing process. With further progress in the field of room acoustics the development of performance practice was strongly influenced as described by Forsyth (1985) et al. Nowadays, even more interaction exists due to the extensive developments in the field of audio engineering. Several directions of music depend strongly on the developments of electronic musical instruments and dedicated tools for sound design. This work makes a contribution in the field of spatial sound design because the author believes that such will play an important role in the further evolution of sonic arts and music. This thesis introduces the concept of *Spatial Sound Design* based on measured room impulse responses. For the first time, a complete processing chain for spatial sound design including interaction design is described. For this reason, the author has studied the state-of-the-art in each segment and developed a complete system. The processing chain consists of *Wave Field Analysis*, *Room Impulse Response Analysis*, *Interaction* and *Spatial Sound Reproduction*. Beside defining the processing chain, emphasis was laid on investigating the limits of a practical system. This is done by choosing a possible state-of-the-art realization for each required step and analyzing the limits compared to the desired situation.

Acknowledgments

This work has been carried out during my years (2004 to 2009) at the Fraunhofer Institute for Digital Media Technology in Ilmenau, Germany, in strong cooperation with the Laboratory of Acoustical Imaging and Sound Control at Delft University of Technology, the Netherlands. Without the support of many people this thesis had never been possible, therefore I like to thank the following people for their extraordinary support during the last years.

First of all I would like to thank my supervisor Diemer de Vries and promoter Prof. Dries Gisolf for their belief in this work even if the direction changed during the studies, their strong support and fruitful and helpful discussions, and the time to read and comment the different versions of the script. I would like to thank Sandra Brix, Prof. Karlheinz Brandenburg and Thomas Sporer for making this dissertation possible during my work at the Fraunhofer Institute for Digital Media Technology IDMT. Without their support this work would never have been finished. I thank all members of the committee for taking time reading the script and take an active part in the defense of this thesis.

Deep thanks goes to all colleagues at the Delft University of Technology, especially Lars Hörchens for doing the first proof reading and giving me a huge number of helpful comments and corrections on the later version of the script. I thank Anton Schlesinger for bringing the spherical array in my mind and Jasper van Drop Schuitman for fruitful and interesting discussions. Thanks for all the inspiring kitchen experiments and accommodation in Delft. I like to thank the staff of the Acoustic Imaging and Sound Control group for making my work and stay in Delft always easy and effective, especially Margaret van Fessem and Henry den Bok for helping me with all bureaucratic details.

Acknowledgments go to Prof. Bernd Fröhlich, Prof. Hans-Peter Schade, Prof. Heidi Krömker, Prof. Karlheinz Brandenburg, and Giovanni Del Galdo for the successful supervision of several student projects as a part of this work.

Thanks to all my former colleagues at Fraunhofer IDMT, especially Judith Liebetrau for supporting listening experiments and social humanity, Clau-

dia Heinze, Michael Strauss, Daniel Beer, Gabriel Gatzsche, Andreas Franck, Clemens Clausen, Haymo Kutschbach, Thomas Röder, Michael Beckinger and Jens-Oliver Fischer.

For the construction and support during the application of the spherical array prototype, I like to thank Stefan Husung and all colleagues of the Institute for Media Technology at the Technical University, Ilmenau.

I like to thank all the students who have worked under my supervision on the various topics combined in this thesis. The chapter on spherical microphones is based on the work of Oliver Thiergart, who has written his Diploma thesis Thiergart (2007) as part of this project and did further investigation in the project *Optimierung akustischer Array-Systeme* founded by the Fraunhofer Gesellschaft and Fraunhofer IDMT. This project was initiated and conducted during my work at Fraunhofer IDMT. The chapter on wave field synthesis is based on the work of Christoph Sladeczek (2008). Furthermore, his work and the work of Michael Rath on the implementation of various prototypes were invaluable for this work and has brought the theoretical ideas to the real world. The chapter on interaction includes results of the student projects of Susanne Gramann (2006), Tobias Laubach (2004), Schneidmadel and Seidenzahl (2005). Many thanks to all other students who contributed different little parts of the whole puzzle.

I like to thank all good friends who have kept me working over the years: Gabriel Gatzsche for his support and always asking the right question, Romy Schnelle who kept me working and see thing so clear, Jan Langhammer for proof reading and fruitful discussion as a friend, while sharing the office at Fraunhofer IDMT and during the last years. I thank Florian Walter for his inspiring talks and support, Beatrix and Stefan Fuhrmann for proof reading the early script. Martin Lambertz for drawing me a hand and an ear. Last but not least I thank my wife Marie-Josefin Melchior for her unlimited love and support and my parents Heinz-Jürgen and Gerlinde Melchior for everything.

Frank Melchior, Erfurt 2011

Contents

1	Introduction	1
1.1	Spatial Sound Reproduction	1
1.2	Spatial Sound Design	2
1.3	Spatial Sound Design Independent of the Reproduction System	4
1.4	Outline of the Thesis	5
2	Fundamentals	7
2.1	Acoustic Wave Equation	7
2.2	Wave Equation in Cartesian Coordinates	11
2.3	Wave Equation in Spherical Coordinates	14
2.4	Kirchhoff-Helmholtz and Rayleigh Integrals	16
2.4.1	Kirchhoff-Helmholtz Integral	17
2.4.2	Rayleigh Integrals	19
2.5	Short Time Fourier Transform	21
2.6	Non-Uniform Rational B-Splines	22
2.7	Correlation and Coherence Measurements	25
2.7.1	Basis Parameter	26
2.7.2	Spatial Correlation and Coherence	27
3	Wave Field Analysis	33
3.1	Introduction	33
3.1.1	Spatial Sampling	34
3.1.2	Virtual Sensor Arrays	36
3.1.3	Beamforming	36
3.1.4	Outline of the Chapter	38
3.2	Spherical Harmonic Decomposition	38
3.3	Discrete Spherical Apertures	43
3.4	Plane Wave Decomposition	49
3.4.1	Spatial Resolution	51
3.4.2	Frequency Bandwidth	55

3.5	Error Analysis of Virtual Spherical Sensor Apertures	57
3.5.1	Description of Spatial Aliasing and Measurement Errors	57
3.5.2	Spatial Aliasing	61
3.5.3	Microphone Noise	63
3.5.4	Positioning Errors of the Sensor Array	66
3.5.5	Non-ideal Microphone Characteristics	71
3.6	Bandwidth Extension Using Multi-radii Design	76
3.6.1	Comparison of Multi-radii Designs	79
3.7	Quantitative Analysis of Measurement Errors	80
3.8	Summary	87
4	Room Impulse Response Analysis	89
4.1	Introduction	89
4.1.1	Outline of the Chapter	90
4.2	Basic Processing Chain	90
4.3	Measurement and System Equalization	92
4.4	Segmentation	94
4.5	Detection and Extraction of Reflections	97
4.5.1	Detection of Reflections	99
4.5.2	Extraction of Reflections	99
4.5.3	Extraction of Reflections Based on Measurements	104
4.6	De-noising of the Diffuse Part	107
4.7	Archiving of Array Measurements	109
4.8	Summary	110
5	Wave Field Extrapolation	111
5.1	Introduction	111
5.1.1	Outline of the Chapter	112
5.2	Plane Wave Extrapolation	112
5.3	Spherical Harmonics Extrapolation	113
5.4	Identity of Plane Wave and Spherical Harmonic Extrapolation	115
5.5	Size of the Extrapolation Area and Extrapolation Errors	117
5.6	Coherence in Extrapolated Fields	120
5.6.1	Influence of the Extrapolation Distance	121
5.6.2	Influence of Spatial Aliasing	121
5.7	Summary	126

6	Interaction	127
6.1	Introduction	127
6.1.1	Basic Chain of Interaction	128
6.1.2	Outline of the Chapter	129
6.2	Taxonomy of Visualization	129
6.2.1	Visualizations of Single Impulse Responses	131
6.2.2	Three-axis Diagrams	133
6.2.3	Visualization of Multiple Impulse Responses	136
6.3	Static Interaction	143
6.3.1	Shaping Surfaces	144
6.3.2	Interaction with Single Impulse Responses	145
6.3.3	Interaction with Direction-dependent Impulse Responses	150
6.4	Dynamic Interaction	153
6.4.1	Source Signal	154
6.4.2	Source Position	154
6.4.3	Source Directivity	156
6.4.4	Signal Processing Depending on Source Parameter	157
6.5	User Interfaces	160
6.5.1	Standard Interface for Spatial Sound Design	163
6.5.2	Extended Interface for Spatial Sound Design	166
6.6	Summary	170
 7	 Spatial Audio Reproduction	 173
7.1	Introduction	173
7.1.1	Outline of the Chapter	174
7.2	Wave Field Synthesis	174
7.2.1	WFS Synthesis Operator	175
7.2.2	Single-user Optimized Wave Field Synthesis	180
7.2.3	Reproduction Data for WFS Reproduction	185
7.2.4	Dynamic Interaction	187
7.3	Stereophonic Reproduction	188
7.3.1	Stereophonic Microphone Setups	188
7.3.2	Principles of Virtual Microphones	192
7.3.3	Reproduction Data for Virtual Room Microphones	193
7.3.4	Dynamic Interaction	193
7.4	Binaural Reproduction	196
7.4.1	Reproduction Data for Binaural Reproduction	199
7.4.2	Dynamic Interaction	201
7.4.3	Binaural Reproduction Using Head-tracking	202

7.5	Summary	202
8	Perceptual Evaluation	203
8.1	Introduction	203
8.2	Influence of Correct Reproduction of Early Reflections	204
8.2.1	Perception of Source Direction	205
8.2.2	Listening Experiment	205
8.2.3	Results	207
8.2.4	Perception of Source Position	208
8.3	Verification of Single-user WFS Optimization	212
8.3.1	Optimization for Focused Sources	212
8.3.2	Diffuse Field Reproduction	214
8.4	Perception of Virtual Microphones	216
8.4.1	Listening Experiment with Mirror Image Source Model	217
8.4.2	Listening Experiment Diffuse Field	221
8.5	Perception of Binaural Reproduction	227
8.5.1	General Design of the Listening Experiments	228
8.5.2	Listening Experiment Mirror Image Source Model	228
8.5.3	Listening Experiment Diffuse-field Simulation	230
8.5.4	Listening Experiment Using a Plane Wave Simulation	230
8.6	Summary	233
9	Conclusions and Future Work	235
9.1	Conclusions	235
9.2	Future Work	236
9.3	Applications of Results	237
A	Mathematic Formulations	239
A.1	Coordinate Systems	239
A.2	Fourier transformations	240
A.2.1	Fourier Transform	240
A.2.2	Multidimensional Fourier Transformation	241
A.2.3	Fourier Series	242
A.2.4	Discrete Fourier Transform	242
A.3	Special Functions	243
A.3.1	Legendre Functions	243
A.3.2	Spherical Bessel and Hankel Functions	243
A.3.3	Spherical Harmonics	244
A.4	Air Absorption	246

A.5	Coherence of Arbitrary Stereophonic Microphone Setups	248
B	Additional Plots	253
B.1	Plots Specific to Section 3.5.2	253
B.2	Plots Specific to Section 3.5.3	255
B.3	Plots Specific to Section 3.5.4	257
B.4	Plots Specific to Section 3.5.5	261
C	Glossary of Acronyms	263
	Bibliography	265
	Summary	283
	Samenvatting	287
	Curriculum Vitae	291

Chapter 1

Introduction

1.1 Spatial Sound Reproduction

Developments in the area of spatial sound reproduction have led to a large variety of established audio systems. Besides experimental set-ups used by the acoustic avantgarde especially in the area of electronic music (Baalman, 2008), several systems are used in the area of movie sound reproduction (Yewdall, 1999) (Rumsey, 2001). This industry is a driving force to bring spatial audio to a large audience. Several development lines and approaches for the reproduction of spatial audio can be found.

A closer look at current developments shows a strong convergence in several systems. On the one hand, Stereophonic Systems are extended and growing from two channels via the ITU-R BS.775 surround setup (Rumsey, 2001) to larger systems like 22.2 (Hamasaki et al., 2005). On the other hand, sound field reproduction systems aiming to reconstruct an acoustic field like Wave Field Synthesis (WFS) (Berkhout et al., 1993) and Ambisonics (Gerzon, 1973), are on the verge of being available on the market. The convergence of loudspeaker based reproduction systems (aiming to reproduce a sound field) is an actual point of research and discussion within the scientific community as started by Nicol (1999) and more recently investigated by Ahrens (2010). Additionally, binaural reproduction is established especially for simulation and auralization applications and psychoacoustic research. Furthermore, there are several reproduction systems trying to combine the advantages of different systems (e.g., Binaural Sky (Menzel et al., 2005), Binaural Room Scanning (Mackensen et al., 1999), WFS with Vector Based Amplitude Panning (VBAP) (Pulkki, 1997)).

With spatial reproduction systems only very few applications are aiming for a natural reproduction of a recorded situation. In most cases the aim is to

communicate artistic messages or ideas:

- A recorded music performance transformed to a spatial reproduction by sound engineers.
- A pure virtual piece of music (e.g., pop music produced in a studio or electronic music).
- A virtual piece of acoustic art (e.g., radio drama).
- An audio-visual artwork (e.g., a movie and its corresponding sound track).

Most applications do not reproduce a real acoustic environment. The spatial audio scene is a pure virtual construct. The development and realization of such a scene is termed *sound design*. The sound designer tries to communicate an acoustical idea and needs to transform this abstract concept into acoustic reality in a given environment with a given reproduction system. Such a concept of sound is not necessarily described by the use of physical models in terms of geometrical room models with an arrangement of real sound sources. Furthermore, such an acoustic idea does not and should not depend on a particular reproduction system.

1.2 Spatial Sound Design

As explained in the previous section, many applications use a spatial sound reproduction system to reproduce a designed virtual acoustic scene.

During the last decades of audio signal processing development, a countless number of tools for the modification of single audio streams¹ have been developed (e.g., Equalizer, Compressor, Modulation effects). All these tools can be used to modify a property of an audio stream. The sound designer transforms his acoustic idea into a parameter set for processing devices to reach his goals of acoustic communication. Beside the artistic knowledge, a strong background in signal processing and the interaction of both is required. Especially, the perceptual effect of a modification of a property of an audio stream is the key element in the know-how of a sound engineer, sound designer or Tonmeister.

In contrast, the process of spatial sound design² modifies the spatial properties

¹The term audio stream is used for a single audio signal in the digital or analog domain.

²The term 'spatial sound design' was used in a different meaning by Herder and Novotny (2003) and Herder (1998) in the context of spatial sound reproduction in virtual environ-

of an audio stream including its position, direction, orientation in a virtual room and the room itself. The simulation of an acoustic environment is one aspect, but also its direction-dependent visualization and modification by the user (e.g., a direction-dependent editing of the early reflections of a virtual source).

Spatial Sound Design is defined as the direction-dependent application of signal processing to a single audio stream and its spatial reproduction.

In comparison to the variety of tools that process a single stream, the number and possibilities of spatial sound design tools have been very limited till now. In current systems the spatial layout of an acoustic scene becomes more important, but the tools for spatial sound design are limited. Often the common tools are specific to a single reproduction system. Furthermore, the effort is concentrated on the task of direct sound positioning (Meltzer et al., 2008) or on integrating the simulation of distance (Hamasaki et al., 2004).

For the sound design process, there are two possible models:

1. The virtual acoustic scene is referred to as an object-oriented virtual reality composed of sound objects and sound manipulating objects (e.g. walls). An acoustic environment as it can be found in the physical world is modeled.
2. The virtual acoustic scene and corresponding acoustic field are visualized³ (direction-dependent) in terms of direction-dependent perceptual or physical properties, but without a representation of physical possible objects or sound sources.

The first model limits the sound designer by physical constraints, which are part of the scene description and have to be implemented on basis of simulations. Moreover, the sound designer has to adapt his acoustic idea to simulated objects. A more intuitive way to modify the sound field is a direct interaction with a graphical representation as given in the latter approach.

The main objective of this thesis is to develop a spatial sound design system, which is not bounded by descriptions of physical and geometrical room acoustics and of which the interaction principles are independent of the reproduction system. The focus of this work lies on the sound design in terms of

ments.

³Visualization of a physical properties along the room dimensions e.g. sound pressure snapshot or impulse responses.

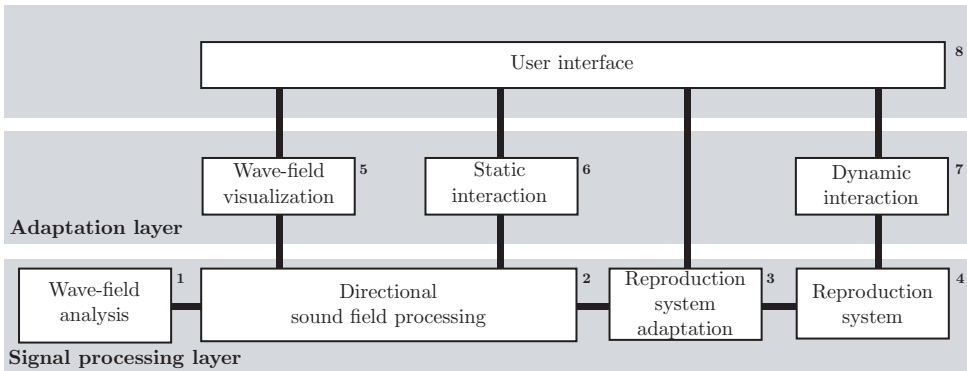


Figure 1.1: Basic block diagram of the system developed in this thesis.

acoustic environments, but the introduced principles can easily be extended. It is important to notice that the basic material for sound design is the analysis of existing rooms. Using the principles of *Spatial Sound Design* these can be modified. At this point a geometrical or physical correspondence is not required.

1.3 Spatial Sound Design Independent of the Reproduction System

For the derivation of spatial sound design principles that are independent of the reproduction system, the following boundary conditions are assumed:

- An acoustic field can be represented as a superposition of plane waves in the far field of its sources.
- A room can be treated as a linear time-invariant (LTI) system so that it can be fully characterized by its impulse response (IR).

As a result, the sound field in a given point can be fully described by a plane wave decomposition of its room impulse response (RIR). Furthermore, based on a plane wave decomposition, the sound field can easily be extrapolated to another point as well as adapted to a specific reproduction system. This adaptation is also part of the sound design process. These assumptions lead to a general system shown in Figure 1.1. The block diagram also represents the structure of this thesis. An acoustic environment has to be captured and ana-

lyzed to serve as a basis for the directional sound field processing. For interaction purposes, the acoustic scene is represented as a plane wave decomposition from the perspective of the sound designer. This representation is visualized. For the visualization step it is important to extract and present the information that is correlated to the perception and avoid *redundant* information. The user can take advantage of several tools to modify the graphical representation in the interaction process. In this step it is important to make changes directly audible and to make the interaction process as efficient as possible. The modified plane wave decomposition is adapted to the target reproduction system using parameters specified by the sound designer. The result of the adaptation process is system specific data required for reproduction. During the reproduction a user can modify several parameters of the acoustic scene in a dynamic interaction process.

1.4 Outline of the Thesis

The mathematical and physical basics required for the developments in the subsequent chapters are briefly reviewed in Chapter 2. In Chapter 3, methods and analysis techniques to achieve a plane wave decomposition based on spherical microphone array measurement data are studied. After a short review of alternative analysis techniques, the cardioid open sphere spherical array is studied in detail. The influence of various measurement errors on the result of the measurements is investigated. The application of spherical microphone arrays to room impulse response analysis is discussed in Chapter 4. The required task of segmentation and extraction of single events from the measured field is discussed. The result of such analysis delivers the basis for interaction and visualization. A method for archiving the measurements is also developed. The extrapolation of measured acoustic fields and its limitations are studied in Chapter 5. Chapters 3 to 5 corresponds to block (1) and (2) in Figure 1.1. The adaptation layer will be treated in Chapter 6. This includes the visualization techniques (5) as well as interaction (6, 7) and the user interface (8) of such a system. A taxonomy of impulse response visualizations is developed and extended by new visualization techniques for the direction-dependent visualizations. Based on these visualizations the concept of shaping surfaces is introduced and used for editing of time variant filtering and spatial editing. The classification into static interaction and dynamic interaction will be given and all steps of the interaction process are studied in detail. A new user interface

concept will be introduced and a prototype system based on conventional user interface elements is described. The adaptation to three different reproduction systems (3) (4) is presented in Chapter 7. The adaptations are developed in view of single user reproduction. While this is typical for binaural reproduction and stereophonic reproduction, a new approach for user-dependent optimization of wave field synthesis will be developed. The generation of reproduction data and the dynamic interaction process is studied from a signal processing point of view.

In Chapter 8 special attention is given to the influence of objective errors on the perceived quality. For wave field synthesis, the concept of user-dependent reproduction is developed and verified in a listening experiment. For binaural reproduction and stereophonic reproduction the influence of measurement errors in spherical array systems is evaluated.

Chapter 9 gives a summary of the complete work and its contributions. Future work is outlined from the author's perspective and a future vision of spatial sound design concludes this thesis.

Chapter 2

Fundamentals

This chapter gives an overview of the mathematical as well as the general fundamentals and definitions required for the subsequent chapters of this thesis. Sections 2.1 to 2.3 summarize the derivation of the wave equation and its specific solutions in different coordinate systems. The Kirchhoff-Helmholtz integral is reviewed in Section 2.4. The short time Fourier transform (STFT) is reviewed in Section 2.5. The description of parametric surfaces as non-uniform rational B-Splines is given in Section 2.6. The measurement of correlation and coherence is introduced in Section 2.7.

2.1 Acoustic Wave Equation

This section follows (Williams, 1999), (Hulsebos, 2004) and (Spors, 2005). For the derivation of the homogeneous acoustic wave equation¹, the following assumptions are made:

1. In the entire region under consideration, the physical properties (e.g., density, temperature) of the medium are independent of the position: the medium is a homogeneous propagation medium.
2. A quiescent propagation medium is assumed, which assures a time independence without movement of the medium.
3. An ideal gas as propagation medium enables the application of the laws for an ideal gas.

¹The wave equation describes the propagation of acoustic pressure or particle velocity as a function of space and time in a medium. The case of a source-free region is covered by the homogeneous wave equation.

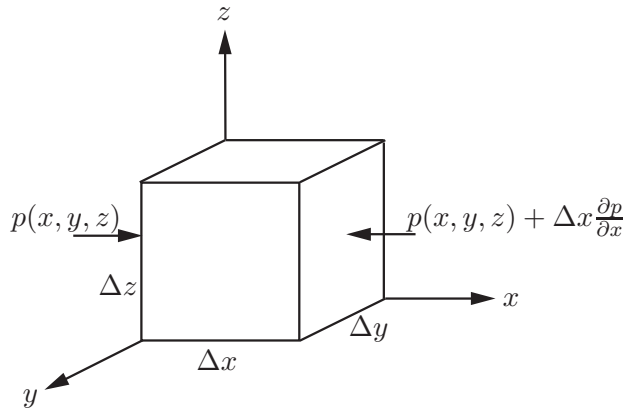


Figure 2.1: Infinitesimal volume element used for the derivation of Euler's equation.

4. State changes in the gas are modeled as adiabatic processes.
5. The perturbations of pressure and density due to wave propagation are small compared to the static pressure p_o and the static density ρ_o . As a result the field variables and medium characteristics can be linearized around an operating point.

A detailed discussion of these assumptions can be found in (Möser, 2007), (Giron, 1996) based on the work of Pierce (1981), or (Morse and Ingard, 1968). The assumptions are appropriate for scenarios in which acoustic wave propagation in rooms and air is considered. Two fundamental physical principles are used to derive the wave equation: the momentum equation and the conservation of mass. The momentum equation relates the force applied to an infinitesimal volume element to the acceleration of this volume element as a result of the force. For the derivation of Euler's equation, first an infinitesimal volume element of fluid $\Delta x \Delta y \Delta z$ as shown in Figure 2.1 is assumed. Due to the pressure $p(x, y, z)$ in the medium all six faces will experience forces. Pressure is a scalar quantity of the unit of force per unit area [N/m^2] or Pascal [Pa]. Positive values correspond to a compression while negative values correspond to expansion. If the pressure at the left face of the volume element in Figure 2.1 is positive, a force will be exerted in the positive x direction. The magnitude of this force is given by $p(x, y, z) \Delta y \Delta z$. The pressure on the opposite face $p(x + \Delta x, y, z)$ is exerted in the negative x direction. The force on the right

hand side is expanded in a Taylor series up to the first order:

$$p(x, y, z) + \Delta x \frac{\partial p}{\partial x}. \quad (2.1)$$

The total force exerted on the volume in the x direction can then be given as

$$F_x = [p(x, y, z) - p(x + \Delta x, y, z)] \Delta y \Delta z = -\Delta x \Delta y \Delta z \frac{\partial p}{\partial x}. \quad (2.2)$$

where \mathbf{F} is the force with F_x as its x-component. Using Newton's equation

$$F_x = ma_x = m \frac{\partial u}{\partial t}, \quad (2.3)$$

with t denoting the time, u denoting the velocity component in the direction of the unit vector in the x -direction defined by Eq.(2.6) and the mass m is given as $m = \rho_0 \Delta x \Delta y \Delta z$, ρ_0 being the fluid density yielding²

$$\rho_0 \frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x}. \quad (2.4)$$

The combination of the same analysis for the y and z direction using vector notation results in the *Euler equation*:

$$\rho_0 \frac{\partial \mathbf{v}(\mathbf{x}, t)}{\partial t} = -\nabla p(\mathbf{x}, t), \quad (2.5)$$

where $\mathbf{v}(\mathbf{x}, t)$ represents the velocity vector at the position \mathbf{x} in the medium given by

$$\mathbf{v} = ue_x + ve_y + we_z. \quad (2.6)$$

The gradient or nabla operator ∇ is defined as (Bronstein et al., 1995):

$$\nabla \equiv \frac{\partial}{\partial x} \mathbf{e}_x + \frac{\partial}{\partial y} \mathbf{e}_y + \frac{\partial}{\partial z} \mathbf{e}_z. \quad (2.7)$$

²According to assumption (1) it is a constant because the medium is assumed to be homogeneous.

The mathematical formulation of the conservation of mass³ is given by:

$$\frac{\partial \varrho}{\partial t} + \varrho \nabla \mathbf{v}(\mathbf{x}, t) = 0. \quad (2.8)$$

ϱ denotes the density of the propagation medium and $\mathbf{v}(\mathbf{x}, t)$ the particle velocity at the position \mathbf{x} . The relation between the time derivative of the density and the acoustic pressure can be found, if assumptions (3)-(5) given on page 7 are assured (Blackstock, 2000):

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = c^2 \frac{\partial \varrho(t)}{\partial t}. \quad (2.9)$$

In Eq. (2.9) c denotes the speed of sound. In case of the medium air under typical conditions ($T = 20^\circ C$, 50% rel. humidity) $c = 343 \frac{m}{s}$. Eq. (2.9) can be used to eliminate the temporal derivative of the density ϱ in Eq. (2.8). If assumption (5) is valid, the result is

$$- \frac{\partial p(\mathbf{x}, t)}{\partial t} = \varrho_0 c^2 \nabla \mathbf{v}(\mathbf{x}, t). \quad (2.10)$$

Eq. (2.5) and Eq. (2.10) in conjunction with initial and boundary conditions form a complete set of first order linear partial differential equations with a unique solution. They can be converted to a single second order equation. For this purpose the time derivative of Eq. (2.10) is taken and inserted into Eq. (2.5). This results in the homogeneous wave equation given by

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = 0, \quad (2.11)$$

where the Nabla operator is defined as (Bronstein et al., 1995)

$$\Delta = \nabla \cdot \nabla = \nabla^2. \quad (2.12)$$

³For a fixed volume inside a fluid the net mass at any time can be taken as the volume integral of the density representing the local average of mass per unit volume in the vicinity of a spatial point. Conservation of mass requires that the time rate of change of the mass is equal to the net mass per unit time entering the volume through the confining surface (Pierce, 1981).

Eq. (2.11) can be transformed to the frequency domain using a Fourier transform as given in Appendix A.2.1, resulting in:

$$\nabla^2 P(\mathbf{x}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{x}, \omega) = 0. \quad (2.13)$$

This equation is known as the *Helmholtz equation*. The term ω/c is abbreviated as the acoustic wavenumber:

$$k^2 = \left(\frac{\omega}{c}\right)^2. \quad (2.14)$$

Eq. (2.14) gives the connection between the acoustic wavenumber k and the temporal frequency $\omega = 2\pi f$. For this reason, variables depending on the temporal frequency ω can be defined by equations including only the wavenumber $k = k(\omega)$.

Solution of the wave equation The acoustic wave equation describes the propagation for a free-field case of a source free volume. To come to a specific solution, information about the boundaries and the sources are required:

1. Initial conditions
2. Acoustic sources
3. Boundary conditions

Depending on the specific problem (e.g. measurement setup), it is convenient to choose a specific coordinate system. In the next section the wave equation in different coordinate systems will be discussed.

2.2 Wave Equation in Cartesian Coordinates

The homogeneous wave equation derived in Section 2.1 can be formulated in Cartesian coordinates as

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = 0. \quad (2.15)$$

The Nabla operator expressed in Cartesian coordinates is

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (2.16)$$

A solution of the homogeneous wave equation will be obtained by separation of variables (Williams, 1999). The tempo-spatial dependence of the pressure $p(\mathbf{x}, t)$ can be separated into individual functions of the different spatial coordinate axes and the time:

$$p(\mathbf{x}, t) = p(x, y, z, t) = p_x(x)p_y(y)p_z(z)p_t(t). \quad (2.17)$$

The solution of the Helmholtz equation for the separated function is by (Ziomek, 1995)

$$p(x, \omega) = A_x(\omega)e^{-ik_x x} + B_x(\omega)e^{+ik_x x}, \quad (2.18a)$$

$$p(y, \omega) = A_y(\omega)e^{-ik_y y} + B_y(\omega)e^{+ik_y y}, \quad (2.18b)$$

$$p(z, \omega) = A_z(\omega)e^{-ik_z z} + B_z(\omega)e^{+ik_z z}, \quad (2.18c)$$

where $A_{\langle \cdot \rangle}(\omega)$ and $B_{\langle \cdot \rangle}(\omega)$ represent arbitrary constants to be solved on the boundary conditions. By applying vector notation as given in Appendix A.1, the result is

$$P(\mathbf{x}, \omega) = A(\omega)e^{-i\mathbf{k}\mathbf{x}} + B(\omega)e^{i\mathbf{k}\mathbf{x}}. \quad (2.19)$$

Eq. (2.19) satisfies the homogeneous wave equation in the frequency domain as long as

$$k^2 = k_x^2 + k_y^2 + k_z^2. \quad (2.20)$$

Since k is constant, the three wavenumber components are not independent of each other (Williams, 1999). A maximum of two can be chosen as independent variables, the third one is dependent. For the two independent variables there are no restrictions, they can extend over all real numbers. It is important to notice that Eq. (2.20) also includes the case of evanescent waves. Evanescent waves have relevance in the studies of radiation from plates and wave reflection and transmission between two different media. Since, only the far field of sources and boundaries is investigated in this work, the reader is referred to (Williams, 1999) for a detailed treatment of evanescent waves. Eq. (2.19) represents a plane wave solution to the wave equation. Since wave propagation in free space is considered here, meaning wave propagation in absence on boundaries, reflected waves do not exist. Therefore, if a plane wave is initially traveling in the positive x, y, z direction there will be no reflected wave traveling

in the negative x, y, z direction resulting in (Ziomek, 1995):

$$B(\omega) = 0. \tag{2.21}$$

To analyze the time dependence in Eq. (2.19), the constant in Eq. (2.19) is set to

$$A(\omega) = 2\pi C(\omega)\delta(\omega - \omega_0), \tag{2.22}$$

where $C(\omega)$ represents an arbitrary constant, hence

$$P(\mathbf{x}, \omega) = 2\pi C(\omega)\delta(\omega - \omega_0)e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{2.23}$$

To arrive at the plane wave in the time domain, the inverse temporal Fourier transform given in Appendix A.2.1 of Eq. (2.23) is taken

$$p(\mathbf{x}, t) = Ae^{i(\mathbf{k}\cdot\mathbf{x} - \omega_0 t)}. \tag{2.24}$$

This solution of the the wave equation is also known as a form of the solution of d'Alembert (Pierce, 1981) given in general as

$$p(\mathbf{x}, t) = f(ct - \mathbf{n} \cdot \mathbf{x}), \tag{2.25}$$

where $f(\cdot)$ denotes an arbitrary function and \mathbf{n} a normal vector indicating the direction of the plane wave (normal to the plane of constant value). It can be related to the wave vector with

$$\mathbf{k} = \frac{\omega}{c}\mathbf{n}. \tag{2.26}$$

The vector \mathbf{k} will be denoted as the wave vector of a plane wave. The acoustic dispersion relation Eq. (2.14) relates the acoustic wave number k to the frequency ω . In view of Eq. (2.26), the length of the wave vector is related to a specific frequency. As a result, each wave vector belongs to a specific frequency ω_0 (Spors, 2005). A signal $e^{i\omega_0 t}$ is called monochromatic. The angular frequency is related to the frequency f_0 by

$$f_0 = \frac{\omega_0}{2\pi}, \tag{2.27}$$

where f_0 is the number of cycles per second the signal exhibits given in Hertz Hz. A plane-wave can be represented using the condensed notation (Williams, 1999)

$$P(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{2.28}$$

Plane waves with arbitrary spectrum can be generated by a superposition of monochromatic plane waves with different frequency dependent weights, but with the same direction of the wave vector. It is important to note that the magnitude of a monochromatic plane wave does not depend on the coordinates in space (see Figure 2.2 for a monochromatic plane wave). Furthermore, it can be shown that for plane waves the pressure p and the particle velocity \mathbf{v} differ only in their absolute values and have always an equal phase (Williams, 1999). A detailed discussion on plane waves including the special case of evanescent waves can be found in (Williams, 1999) and (Ziomek, 1995).

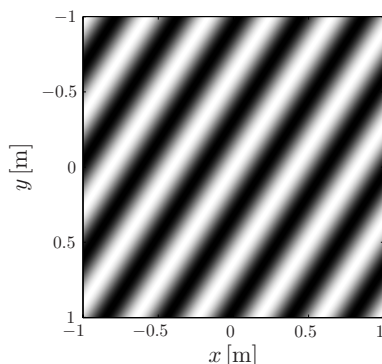


Figure 2.2: Pressure field of a $f_0 = 1$ kHz plane wave in the xy -plane. The gray level illustrates the amplitude.

Plane-Wave Expansion The homogeneous wave equation is fulfilled by the d’Alemberts solution for all possible choices of \mathbf{n} . As a result, arbitrary solutions of the wave equation can be expressed as a superposition of plane waves traveling into all possible directions. If the frequency domain formulation of a plane wave is used, an arbitrary wave field can be expressed as a superposition of all possible wave vectors (Spors, 2005).

2.3 Wave Equation in Spherical Coordinates

In this section, the solution of the wave equation using spherical coordinates as defined in Appendix A.1 is considered. The homogeneous wave equation is

given by

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = 0, \quad (2.29)$$

where \mathbf{x} denotes a point in a spherical coordinate system defined in Appendix A.1. The gradient of the pressure in spherical coordinates is given by

$$\nabla p = \mathbf{e}_r \frac{\partial p}{\partial r} + \mathbf{e}_\theta \frac{\partial p}{\partial \theta} + \mathbf{e}_\phi \frac{1}{r \sin \theta} \frac{\partial p}{\partial \phi} \quad (2.30)$$

This leads to the time dependent wave equation in spherical coordinates

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial p}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 p}{\partial \phi^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \quad (2.31)$$

General solution of the wave equation The solution of Eq.(2.31) can be derived by separation of variables. For the acoustic wave equation in the frequency-domain

$$\nabla^2 P(\mathbf{x}, \omega) + k^2 P(\mathbf{x}, \omega) = 0, \quad (2.32)$$

the detailed derivation can be found in (Williams, 1999, p.185) which leads to the following solutions:

$$P(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left[C_{nm}(\omega) h_n^{(1)}(kr) + D_{nm}(\omega) h_n^{(2)}(kr) \right] Y_n^m(\theta, \phi). \quad (2.33)$$

The *level* n and the *mode* m are integer values with $0 \leq n \leq \infty$ and $-n \leq m \leq n$. The spherical Hankel functions of the first and second kind are given by $h_n^{(1)}(kr)$ and $h_n^{(2)}(kr)$, respectively. The function $Y_n^m(\theta, \phi)$ defines the spherical harmonic level n and model m with

$$Y_n^m(\theta, \phi) \equiv \sqrt{\frac{(2n+n)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \theta) e^{im\phi}. \quad (2.34)$$

$P_n^m(\cos \theta)$ denotes the Legendre function of the first kind. If the time dependence is defined to be $e^{-i\omega t}$, the spherical Hankel function of the first kind $h_n^{(1)}$ corresponds to the incoming spherical waves while the spherical Hankel function of the second kind $h_n^{(2)}$ corresponds to outgoing spherical waves with respect to the origin. By expressing the spherical Hankel functions in spherical Bessel function $j_n(kr)$ and spherical Neumann function $y_n(kr)$ according to

Eq. (A.17), Eq. (2.33) can be written as

$$P(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [A_{nm}(\omega)j_n(kr) + B_{nm}(\omega)y_n(kr)] Y_n^m(\theta, \phi), \quad (2.35)$$

where

$$A_{nm}(\omega) = C_{nm}(\omega) + D_{nm}(\omega), \quad (2.36)$$

$$B_{nm}(\omega) = i(C_{nm}(\omega) - D_{nm}(\omega)), \quad (2.37)$$

The terms $A_{nm}(\omega)$, $B_{nm}(\omega)$, $C_{nm}(\omega)$ and $D_{nm}(\omega)$ are defined by the boundary values of the problem. For an *interior* problem (see Figure 2.3(a)) in which all sources are located outside a sphere of a radius b , the solution consists of radial functions which are finite at the origin O . Since all systems under consideration are physical systems and both Hankel functions and Neumann functions are infinite at the origin, it is required that $B_{nm}(\omega) = 0$. As a result the internal pressure field $P(r, \theta, \phi, \omega) = P(\mathbf{x}, \omega)$ is defined as

$$P(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{mn}(\omega)j_n(kr)Y_n^m(\theta, \phi). \quad (2.38)$$

If the spherical surface completely encloses all sources (see Figure 2.3(b)), the external pressure field is determined by Eq. (2.33). The general solution can be reduced to

$$P(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [C_{nm}(\omega)h_n^{(1)}(kr)] Y_n^m(\theta, \phi), \quad (2.39)$$

because $D_{nm}(\omega) = 0$ since incoming waves cannot exist in this case.

2.4 Kirchhoff-Helmholtz and Rayleigh Integrals

The Kirchhoff-Helmholtz integral forms the basis for the concept of wave field synthesis (WFS). It is a strict mathematical formulation of the principle of Huygens, which was formulated in 1690. Huygens stated that each volume element of a wave front can be regarded as an individual spherical source contribution to the sound wave front at a later time. The position of the full wave front at any later time is the envelope of all such elementary wave fronts. In 1818 the principle was supplemented by Fresnel with the idea that

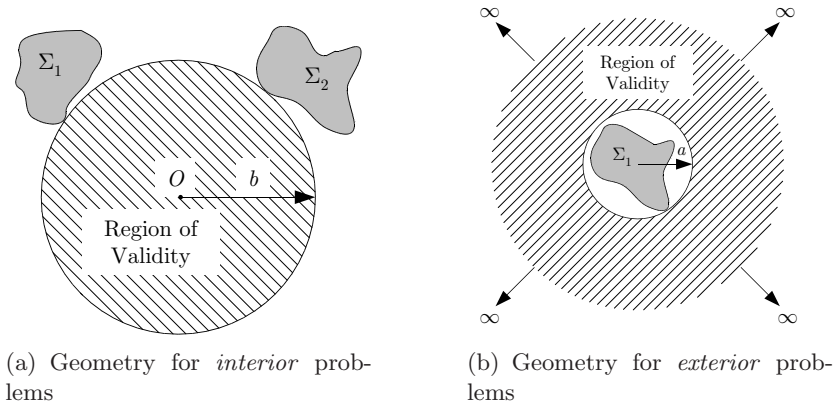


Figure 2.3: Geometry for the definition of source free regions. Source distributions are indicated as Σ .

the secondary wave fronts interfere. This combination is called the Huygens-Fresnel principle. The mathematical basis of this idea was given by Kirchhoff in 1882. He showed that the Huygens-Fresnel principle can be regarded as a special case of a more general theorem. This theorem states that the sound pressure in a source free volume V can be calculated if both sound pressure and normal particle velocity are known on the boundary surface S surrounding V . If the surface is an infinitely large plane separating the area of sources and the source-free region, the Kirchhoff theorem can be simplified to the Rayleigh theorems, stating that the acoustic field on the source free side can be completely reconstructed by means of a continuous distribution of monopoles or dipoles.

2.4.1 Kirchhoff-Helmholtz Integral

The geometry for deriving the Kirchhoff-Helmholtz integral is drawn in Figure 2.4. Inside an unbounded homogeneous medium with the source distribution Σ lies the volume V with the inward pointing unit normal vector \mathbf{n} . The wave field $P(\mathbf{r}, \omega)$ inside the volume V due to the source distribution outside fullfills the homogeneous wave equation given by Eq. (2.13):

$$\nabla^2 P(\mathbf{r}, \omega) + k^2 P(\mathbf{r}, \omega) = 0 \quad (2.40)$$

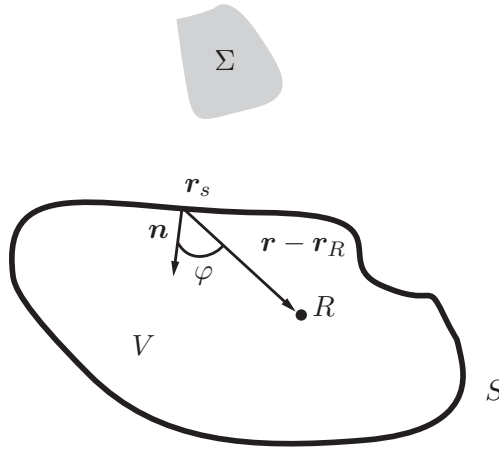


Figure 2.4: Geometry used for the derivation of the Kirchhoff-Helmholtz integral. The source free volume V bounded by the surface S , surrounded by the source distribution Σ .

The second theorem of Green can be applied to the given geometry:

$$\int_V f \nabla^2 g - g \nabla^2 f dV = \int_S g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} dS, \quad (2.41)$$

where f and g are two functions having continuous first and second partial derivatives within and on S . If f and g also satisfy the homogeneous wave equation, the first term vanishes (Blackstock, 2000) and Eq.(2.41) reduces to

$$\int_S g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} dS = 0. \quad (2.42)$$

Now a specific solution of the Helmholtz equation is considered for g , i.e., the pressure field of a monopole source at point R given in the space-frequency domain with a frequency-independent source spectrum $S(\omega) = 1$,

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = \frac{e^{-ik|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \quad (2.43)$$

This function has the requisite continuity properties, except at the point R itself. Furthermore, it fulfills the inhomogeneous wave equation

$$\nabla^2 G(\mathbf{r}|\mathbf{r}_R, \omega) + k^2 G(\mathbf{r}|\mathbf{r}_R, \omega) = -4\pi\delta(\mathbf{r}-\mathbf{r}_R) \quad (2.44)$$

Eq. 2.43 is considered next for a special case of g and f using $G(\mathbf{r}|\mathbf{r}_R, \omega)$ and $P(\mathbf{r}, \omega)$, respectively. The substitution and integration excluding the point R and utilizing the equation of conservation of momentum yields:

$$P(\mathbf{r}, \omega) = \frac{1}{4\pi} \int_S \left[ik\varrho_0 V_n(\mathbf{r}_s, \omega) + P(\mathbf{r}_s, \omega) \frac{1 + ik|\mathbf{r} - \mathbf{r}_R|}{|\mathbf{r} - \mathbf{r}_R|} \cos(\varphi) \right] \frac{e^{-ik|\mathbf{r} - \mathbf{r}_R|}}{|\mathbf{r} - \mathbf{r}_R|} dS \quad (2.45)$$

The Kirchhoff-Helmholtz integral (Eq. (2.45)) states that in a volume the sound pressure at any point can be calculated if pressure and normal particle velocity on the boundary surface are known. This forms the basis for wave field synthesis (WFS) as discussed in detail in Section 7.2.

2.4.2 Rayleigh Integrals

In the derivation of the Kirchhoff-Helmholtz integral the choice of the Green's function is not unique. To simplify the Kirchhoff-Helmholtz integral different Green's functions can be chosen such that a more convenient boundary condition is satisfied. The following two boundary conditions are considered:

$$\nabla_S G(\mathbf{r}|\mathbf{r}_R, \omega) \cdot \mathbf{n} = 0, \quad (2.46)$$

$$G(\mathbf{r}|\mathbf{r}_R, \omega) = 0, \quad (2.47)$$

with \mathbf{r} on S . The condition in Eq. (2.46) implies that S acts as a perfectly reflecting, rigid surface. Under the conditions in Eq. (2.47) S behaves like a perfect pressure-free surface. For the specific geometry given in Figure 2.5, simple solutions can be found for the Green's functions that satisfy these boundary conditions. The reconstruction volume is enclosed by a plane surface S_0 and a spherical surface S_1 . The sources of the acoustical pressure field are situated in the upper half space separated by the plane surface S_0 . After applying Sommerfeld's radiation condition to the boundary and letting r_1 tend to infinity, only surface S_0 will contribute to the Kirchhoff-Helmholtz integral. In other words, only S_0 has to be considered to calculate the pressure in R using the Kirchhoff-Helmholtz Integral given in Eq. 2.45. A Green's function can be found for the geometry in Figure 2.5, such that either the first or the second condition is fulfilled given in Eq. 2.46 and Eq. 2.47, respectively. Since the surface acts like a perfect reflector the Green's function can be interpreted as being the wave field of two monopole sources situated symmetrically with respect to the surface S_0 (Start, 1997). In case of condition Eq. (2.46) (rigid boundary)

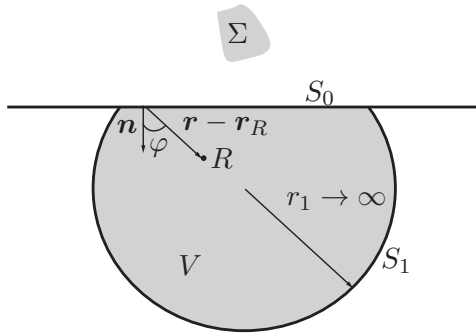


Figure 2.5: Geometry used for the derivation of the Rayleigh integrals. The source free volume V is separated from the source area by the infinite planar surface S_0 .

a suitable Green's function is the sum of two equal phase monopoles situated symmetrically with respect to S_0 . Using this Green's function the Kirchhoff-Helmholtz integral equation reduces to

$$P(\mathbf{r}_R, \omega) = -\frac{1}{4\pi} \int_{S_0} [G(\mathbf{r}|\mathbf{r}_R, \omega) \nabla_S P(\mathbf{r})] \cdot \mathbf{n} dS_0, \quad (2.48)$$

using the equation of conservation of the momentum Eq. 2.5 the integral can be written as

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left[ik \varrho_0 V_n(\mathbf{r}, \omega) \frac{e^{-ik|\mathbf{r}-\mathbf{r}_R|}}{|\mathbf{r}-\mathbf{r}_R|} \right] dS_0, \quad (2.49)$$

which is called the Rayleigh I integral representation. It states that a wave field can be reconstructed by means of a continuous distribution of monopoles at the surface S_0 . The strength of each monopole is proportional to the normal component of the particle velocity $V_n(\mathbf{r}, \omega)$ of the incident wave field measured at the position of that monopole. In order to satisfy the pressure-free boundary condition of Eq. 2.47 the two monopoles considered have opposite polarity. In this case the specific Kirchhoff-Helmholtz Integral becomes

$$P(\mathbf{r}_R, \omega) = -\frac{1}{4\pi} \int_{S_0} [P(\mathbf{r}) \nabla_S G(\mathbf{r}|\mathbf{r}_R, \omega)] \cdot \mathbf{n} dS_0, \quad (2.50)$$

resulting in

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{S_0} \left[P(\mathbf{r}, \omega) \frac{1 + ik|\mathbf{r} - \mathbf{r}_R|}{|\mathbf{r} - \mathbf{r}_R|} \cos(\varphi) \frac{e^{-ik|\mathbf{r} - \mathbf{r}_R|}}{|\mathbf{r} - \mathbf{r}_R|} \right] dS_0, \quad (2.51)$$

which is called the Rayleigh II representation integral. It states that any wave field due to sources in the upper half-space of Figure 2.5 can be reconstructed in the other half space by means of a continuous distribution of dipoles at the surface S_0 . The strength of each dipole is given by the pressure $P(\mathbf{r}, \omega)$ of the incident wave field measured at the position of that dipole.

2.5 Short Time Fourier Transform

Room impulse responses are non-periodic and time-variant signals. For this kind of signal the short time Fourier transform (STFT) can be used to analyze, modify and synthesize the time-varying spectrum (Allen and Rabiner, 1977) (Allen, 1977) (Serra, 1989). The STFT is defined as the Fourier transform (FT) of a windowed version of a function as described by Eq. (A.8a) (Băni, 2002):

$$P(\tau, \omega) \equiv STFT \{p(t)\} = \int_{-\infty}^{\infty} p(t)w(t - \tau)e^{-i\omega t} dt, \quad (2.52)$$

where $w(t - \tau)$ denotes the shifted version of the window function $w(t)$. The results of this operation are spectra depending on ω and the temporal window position τ . The magnitude spectra versus time plot of a STFT is termed a spectrogram (Koenig et al., 1946):

$$spectrogram = |P(\tau, \omega)|. \quad (2.53)$$

The inverse short time Fourier transform (ISTFT) can be calculated using:

$$p(t) \equiv STFT^{-1} \{P(\tau, \omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\tau, \omega)e^{i\omega t} d\tau d\omega. \quad (2.54)$$

In the latter sections of this thesis, the discrete STFT is applied. Analogue to the continuous STFT it is based on the DFT given in Eq. (A.12a). A schematic drawing of the process is depicted in Figure 2.6. The definition of the discrete

STFT of a sampled signal $p(n)$ is (Arfib et al., 2007), (Crochiere, 1980):

$$P(s, k) = \sum_{m=-\infty}^{\infty} p(m)w(sR_a - m)e^{-i\frac{2\pi}{N}mk}, \quad (2.55)$$

where $w(sR_a - m)$ denotes the analysis window, $P(s, k)$ represents the time-varying spectrum with time index s and frequency bins $0 \leq k \leq N-1$. At each time instance n the signal is weighted with the window $w(sR_a - m)$. Thus the FFT of the size N can be computed with an finite sum over m . If the window $w(s)$ has a constant overlap-add property at a given hop-size R_a

$$\sum_{s=-\infty}^{\infty} w(sR_a - m) = 1 \quad \forall s \in \mathbb{Z}, \quad (2.56)$$

the sum of $P(s, k)$ equals the whole signal $P(k)$ in the frequency domain. The synthesis algorithm is given by

$$p(n) = \sum_{s=-\infty}^{\infty} w_s(n - sR_s)p_s(n - sR_s) \quad (2.57)$$

with

$$p_s(n) = \frac{1}{N} \sum_{k=0}^{N-1} \left[e^{-i\frac{2\pi}{N}sR_s k} P(n, k) \right] e^{-i\frac{2\pi}{N}nk},$$

stating that the synthesized signal $p(n)$ is derived from the inverse transformed short-time spectra $P(sR_s, k)$ weighted with the synthesis window w_s and summed using the overlap-add procedure given by Eq. (2.57). The STFT is an identity system, as long as the analysis windows overlap and add to one. The parameter settings become relevant if a modification of the input data is desired. This case will be studied in Section 6.3.2.

2.6 Non-Uniform Rational B-Splines

Two common methods for representing surfaces are implicit equations and parametric functions. The implicit description is given by $f(x, y, z) = 0$. In parametric form each of the coordinates of a point on a surface is represented

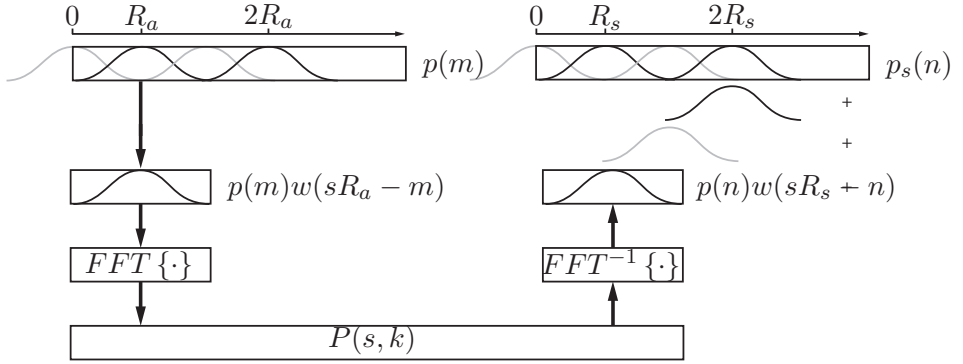


Figure 2.6: Block diagram of the discrete Short Time Fourier Transform. The input signal $p(m)$ is spitted in blocks and windowed. The results are spectra at a discrete time position $P(s, k)$. After processing these spectra are transformed back utilizing a synthesis window and overlapped and added to achieve the new signal $p_s(n)$.

as an explicit function of two independent parameters

$$\mathbf{S}(u, v) = [x(u, v), y(u, v), z(u, v)], \quad (2.58)$$

where $a \leq u \leq b$ and $c \leq v \leq d$. $\mathbf{S}(u, v)$ is a vector-valued function of the independent variables u and v . The intervals $[a, b]$ and $[c, d]$ are usually normalized to $[0, 1]$. As an example, a sphere of unit radius centered at the origin is considered: $x^2 + y^2 + z^2 - 1 = 0$. A parametric representation is given by $\mathbf{S}(u, v) = (x(u, v), y(u, v), z(u, v))$, where

$$x(u, v) = \sin(u) \cos(v), \quad (2.59a)$$

$$y(u, v) = \sin(u) \sin(v), \quad (2.59b)$$

$$z(u, v) = \cos(u), \quad (2.59c)$$

with

$$0 \leq u \leq \pi, \quad 0 \leq v \leq 2\pi \quad (2.60)$$

Several mathematic descriptions exist for parametric surface generation. One with the most widespread acceptance and popularity are Non-Uniform Rational B-Splines (NURBS). The reasons for the use of NURBS in this context are partly the same reasons why they are so widespread in computer aided design (Piegl, 1991):

- NURBS offer a common mathematical form for representation and design of standard analytical shapes and free-form curves and surfaces.
- NURBS provide the flexibility to design a large variety of shapes by manipulating their control points and weights.
- Evaluation is reasonably fast and computationally stable
- Because of their clear geometrical interpretations NURBS are particularly useful for designers.
- Powerful geometric toolkits exist, which can be used to design, analyze, process and interrogate objects.

As described in (Piegl and Tiller, 1997) a NURBS-surface of degree p in u -direction and degree q in v -direction is a bi-variate vector-valued piecewise rational function of the form .

$$\mathbf{S}(u, v) = \frac{\sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u)N_{j,q}(v)w_{i,j}\mathbf{P}_{i,j}}{\sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u)N_{j,q}(v)w_{i,j}}, \quad (2.61)$$

with

$$0 \leq u, v \leq 1 .$$

The $\mathbf{P}_{i,j}$ form a bidirectional control net, the $w_{i,j}$ are the weights, and the $N_{i,p}(u)$ and $N_{j,q}(v)$ are the non-rational B-spline basis functions defined on the knot vectors

$$\begin{aligned} U &= (\underbrace{0, \dots, 0}_{p+1}, u_{p+1}, \dots, u_{r-p-1}, \underbrace{1, \dots, 1}_{p+1}) \\ V &= (\underbrace{0, \dots, 0}_{q+1}, v_{q+1}, \dots, v_{s-q-1}, \underbrace{1, \dots, 1}_{q+1}) \end{aligned} \quad (2.62)$$

where $r = n + p + 1$ and $s = m + q + 1$. The i th B-spline basis function of p -degree (order $p+1$), denoted by $N_{i,p}(u)$, is defined as

$$N_{i,0}(u) = \begin{cases} 1 & \forall u_i \leq u \leq u_{i+1} \\ 0 & \text{otherwise,} \end{cases} \quad (2.63a)$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u). \quad (2.63b)$$

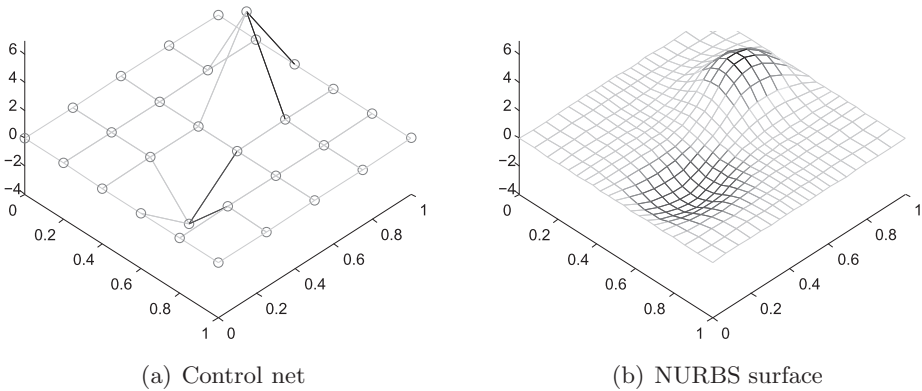


Figure 2.7: (a) Control net $P_{i,j}$ and (b) corresponding NURBS surface $S(u,v)$ for specific weights and knot vectors. The surface was evaluated in a grid of 20 by 20 points.

Figure 2.7 presents the control net $P_{i,j}$ and the corresponding evaluated surface for specific weights and knot vectors. It is important to note that a parametric surface can be evaluated in arbitrary resolutions. This is an important property for the use in spatial sound design applications as given in Chapter 6. The surface can be modified easily by dragging the coefficients of the control net, which correspond to points in the three-dimensional space (i.e., the control net given in Figure 2.7(a)).

NURBS will be used throughout this work as a new interaction tool and representation for two-dimensional parameter editing.

2.7 Correlation and Coherence Measurements

The description of correlation and coherence is important for the description of random data like diffuse sound fields in rooms. In this section the basic parameter will be introduced briefly based on Bendat and Piersol (1980). Furthermore, the spatial coherence of diffuse sound fields will be discussed, following Kuster (2008). A similar analysis can be found in (Elko, 2001) using a different mathematical description.

2.7.1 Basis Parameter

A measure for the linear statistical dependence of two values of a real random variable $x(t)$ at a time lag τ is the autocorrelation function

$$R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t)x(t + \tau)dt. \quad (2.64)$$

The measure of a linear statistical correlation of two real random variables $x(t)$ and $y(t)$ is the cross-correlation function $R_{xy}(\tau)$ given by

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t)y(t + \tau)dt. \quad (2.65)$$

The arithmetic mean $\overline{x(t)}$ of a time dependent variable $x(t)$ is given by

$$\overline{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t)dt. \quad (2.66)$$

The quadratic mean $\overline{x^2(t)}$ is defined by

$$\overline{x^2(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x^2(t)dt. \quad (2.67)$$

Based on the zero mean function $x_z(t) = x(t) - \overline{x(t)}$ the variance $\overline{x_z^2(t)}$ can be calculated using

$$\overline{x_z^2(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x_z(t)^2 dt. \quad (2.68)$$

The square root of the variance is denoted as σ , the standard deviation, given by

$$\sigma = \sqrt{\overline{x_z^2(t)}}. \quad (2.69)$$

The cross-correlation coefficient $\rho_{xy}(\tau)$ is defined as

$$\rho_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}}. \quad (2.70)$$

Insertion of the zero mean functions $x_z(t)$ and $y_z(t)$ in Eq. (2.70) results in

$$\rho_{xy}(\tau) = \frac{\lim_{T \rightarrow \infty} \frac{1}{T} \int x_z(t) y_z(t + \tau) dt}{\sqrt{\left[\lim_{T \rightarrow \infty} \frac{1}{T} \int x_z^2(t) dt \right] \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int y_z^2(t) dt \right]}}. \quad (2.71)$$

Eq. (2.71) gives a general form of the correlation coefficient. Since the correlation coefficient only delivers a single average value, it is a common way to study its frequency dependence by using $1/3$ octave band filters and calculating the value for each frequency band. An alternative approach is to calculate the continuous coherence function based on the cross-spectral density function $S_{xy}(\omega)$ given by

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau} d\tau, \quad (2.72)$$

resulting in

$$\gamma_{xy}^2(\omega) = \frac{|S_{xy}(\omega)|^2}{S_{xx}(\omega) S_{yy}(\omega)}. \quad (2.73)$$

Using Eq. (2.73) a frequency dependent correlation of two signals can be measured. Furthermore, by applying a segmentation in time, a time-frequency correlation diagram can be calculated.

2.7.2 Spatial Correlation and Coherence

In case of signals or room impulse responses measured with a dummy head represented by $x_z(t)$ and $y_z(t)$, the value for $\rho_{xy}(\tau)$ is called inter-aural cross-correlation coefficient (IACC) (ISO 3382, 2000). If the two signals $x_z(\mathbf{x}_1, t)$ and $y_z(\mathbf{x}_2, t)$ represent measurements of sound pressure at different spatial position P_1 and P_2 in a sound field, the spatial correlation coefficient can be calculated. This quantity was first derived by Cook et al. (1955) for two sound pressure measurements. Jacobson (1979) extended it to the coherence function as well as to a derivation for different velocity elements, which is the basis for the alternative derivation of Rafaely (2000) and verified by Chun et al. (2003) and Rafaely (2000). In the context of this thesis the work of Kuster (2008) based on derivations of Jacobson is important, which will be considered in the derivation in Section 7.3.1. The basic idea of the derivation is the

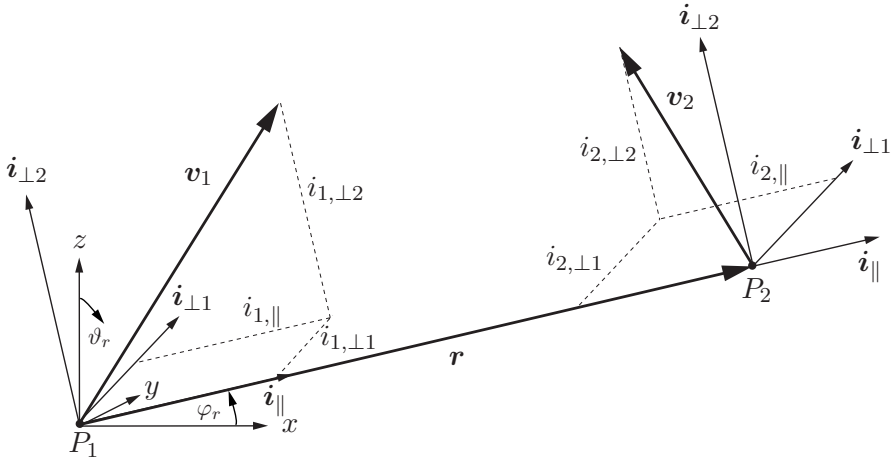


Figure 2.8: Geometry for the decomposition of the velocity vectors \mathbf{v}_1 and \mathbf{v}_2 at two positions denoted by P_1 and P_2 into components in the coordinate system $(\mathbf{i}_{\perp 1}, \mathbf{i}_{\perp 2}, \mathbf{i}_{\parallel})$. Adapted from (Kuster, 2008).

separation of a given first-order directivity into a representation of pressure p and velocity components parallel v_{\parallel} as well as perpendicular v_{\perp} to each other. If two microphone signals are separated in this way, the correlation and coherence functions can be calculated by superposition based on the results derived by Jacobson. Figure 2.8 presents the geometry used in the derivation.

Spatial correlation coefficient As derived by Cook, the spatial correlation coefficient between two pressure signals at a distance r with frequency ω_0 and a time lag of τ is given by

$$\rho_{pp}(r, \tau) = \frac{\sin(kr)}{kr} \cos(\omega_0\tau). \quad (2.74a)$$

As derived in (Jacobson, 1979) the following expressions can be found for the spatial correlation coefficients between the pressure and different components of the velocity.

$$\rho_{pv_{\parallel}}(r, \tau) = \sqrt{3} \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} \cos(\omega_0\tau) \quad (2.74b)$$

$$\rho_{v_{\parallel}v_{\parallel}}(r, \tau) = 3 \frac{(kr)^2 \sin(kr) + 2kr \cos(kr) - \sin(kr)}{(kr)^3} \cos(\omega_0\tau) \quad (2.74c)$$

$$\rho_{v_{\perp}v_{\perp}}(r, \tau) = 3 \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} \cos(\omega_0\tau) \quad (2.74d)$$

The derivation of the correlation coefficient by Jacobsen is based on the correlation function of a diffuse field model with plane waves of a mean square magnitude G^2 , resulting in

$$R_{pp}(r, \tau) = \frac{G^2}{2} \frac{\sin(kr)}{kr} \cos(\omega_0\tau), \quad (2.75a)$$

$$R_{pv_{\parallel}}(r, \tau) = \frac{G^2}{2\rho_0c} \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^2} \sin(\omega_0\tau), \quad (2.75b)$$

$$R_{v_{\parallel}v_{\parallel}}(r, \tau) = \frac{(kr)^2 \sin(kr) + 2kr \cos(kr) - 2 \sin(kr)}{(kr)^3} \frac{G^2}{2(\rho_0c)^2} \cos(\omega_0\tau), \quad (2.75c)$$

$$R_{v_{\perp}v_{\perp}}(r, \tau) = \frac{G^2}{2(\rho_0c)} \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} \cos(\omega_0\tau), \quad (2.75d)$$

$$R_{pp}(0, 0) = \frac{G^2}{2}, \quad (2.75e)$$

$$R_{v_{\parallel}v_{\parallel}}(0, 0) = R_{v_{\perp}v_{\perp}}(0, 0) = \frac{G^2}{6(\rho_0c)^2}. \quad (2.75f)$$

It can be shown that the cross-correlation and correlation coefficients between all other possible combinations of pressure and particle velocity are zero for any distance r (Jacobson, 1979). The following expressions can be derived for the spatial coherence functions between pressure and particle velocity:

$$\gamma_{pp}^2(r, \omega) = \left[\frac{\sin(kr)}{kr} \right]^2, \quad (2.76a)$$

$$\gamma_{pv_{\parallel}}^2(r, \omega) = 3 \left[\frac{\sin(kr) - (kr) \cos(kr)}{(kr)^2} \right]^2, \quad (2.76b)$$

$$\gamma_{v_{\perp}v_{\perp}}^2(r, \omega) = 9 \left[\frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} \right]^2, \quad (2.76c)$$

$$\gamma_{v_{\parallel}v_{\parallel}}^2(r, \omega) = 9 \left[\frac{(kr)^2 \sin(kr) + 2kr \cos(kr) - 2 \sin(kr)}{(kr)^3} \right]^2. \quad (2.76d)$$

Based on these coherence functions, the coherence function for arbitrary first-order sensor elements can be calculated. The result is obtained by superposition of the components given above. These results are useful for characterizing the coherence of spatially separated sensor elements in a diffuse sound field. They deliver objective parameters to compare the simulated results based on spherical array measurements in Chapter 5. A specific discussion and analysis of spatial coherence for stereo microphone setups and their virtual simulations based on spherical array measurements will be given in Section 7.3.

Decorrelation

In the context of room impulse response (RIR) post processing, it is often required to decorrelate⁴ the reverberation tail of an room impulse response. The aim can be to simulate a given or desired coherence function (e.g., to simulate a specific microphone setup) (Muraoka, 2006). Another application is to reduce the degree of correlation especially in lower frequencies. This can be the result of limited spatial resolution in commonly used array sizes. To give a good impression of envelopment and diffusion, low correlation is required. Different techniques for the decorrelation of RIRs can be found in (Merimaa, 2006). (Potard, 2006) studied decorrelation techniques in relation to the reproduction of extended sound sources.

In the context of this work, decorrelation based on the STFT as given in Section 2.5 is used. The aim is to obtain a variable frequency dependent coherence function. For this purpose the time-variant complex spectrum of a signal $P_A(s, k) = DSTFT \{p(n)\}$ is used. Such a signal can be separated into a magnitude and phase spectrum as

$$A(s, k) = 20 \log(|P(s, k)|), \quad (2.77)$$

$$\Theta_A(s, k) = \angle P(s, k). \quad (2.78)$$

Since the decorrelation studied in this context will only be applied to the diffuse part of a room impulse response, it is advantageous to calculate the discrete short time Fourier transform of a sampled white noise signal $w(n)$ given by $W(s, k) = DSTFT \{w(n)\}$. Now the phase spectrum of the noise signal $\Theta_W(s, k)$ can be used to decorrelate the impulse response. A new phase

⁴Reduce the degree of correlation between to signals.

spectrum $\Theta_B(s, k)$ is generated using

$$\Theta_B(s, k) = \Theta_A(s, k) \cdot \gamma(k) + \Theta_W(s, k) \cdot [1 - \gamma(k)], \quad (2.79)$$

where $\gamma(k)$ represents the frequency dependent coherence with $\gamma(k) \in [0, 1]$. The decorrelated room impulse response can be obtained by applying the IDSTFT given by Eq. (2.57). It is important to note that due to the phase manipulation a synthesis window is required. Informal listening experiments by the author have led to good results using a Tukey window (often called a cosine-tapered window). Such a window $w_s(n)$ of width N can be constructed using (Harris, 1978)

$$w_s(n) = \begin{cases} 1, & 0 \leq |n| \leq \alpha \frac{N}{2} \\ \frac{1}{2} \left[1 + \cos \left[\pi \frac{n - \alpha \frac{N}{2}}{2(1-\alpha) \frac{N}{2}} \right] \right], & \alpha \frac{N}{2} \leq |n| \leq \frac{N}{2} \end{cases}. \quad (2.80)$$

The factor α denotes the degree of tapering. The degeneration of the time structure of the impulse response is not critical due to the fact that it is only applied to the diffuse part which can be compared with a noise like signal. Artefacts due to the phase manipulation are reduced to an imperceptible level. An impulse response using this techniques is included in the perceptual experiments in Section 8.4.1.

Chapter 3

Wave Field Analysis Using Spherical Apertures

3.1 Introduction

The basis for spatial sound design is the three-dimensional spatio-temporal decomposition of an acoustic wave field. Such decomposition is enabled by microphone array technology in various resolutions and dimensions. Two-dimensional microphone array designs have been studied by Hulsebos (2004) in the context of auralization for wave field synthesis. The special case of circular arrays has been studied by Spors (2005) in the context of room compensation and by Kuntz (2009) more in detail including a short section on the auralization using binaural reproduction and WFS reproduction. The auralization for pure binaural systems based on spherical arrays has been studied by Duraiswami et al. (2005). For a flexible analysis of acoustic fields, a three-dimensional approach is desirable. Such an analysis can be offered by spherical microphone arrays with a nearly direction-independent resolution. For this reason, this kind of sensor systems are examined for application of wave field analysis in this work. The part described in this chapter corresponds to block (1) in Figure 3.1. The first arrays based on spherical harmonic processing were presented by Meyer and Elko (2002) and Abhayapala and Ward (2002). Recent research has led to new theoretical developments, experimental investigations, and signal processing methods. A brief overview can be found in (Rafaely, 2008a). The following two sections review this summary in the context of this work. Two main aspects have been investigated in recent research: spatial sampling and beam-forming. While the first aspect deals with methods to measure the acoustic field on a limited number of spatial sampling positions, the latter enables the directional analysis of the measurement data.

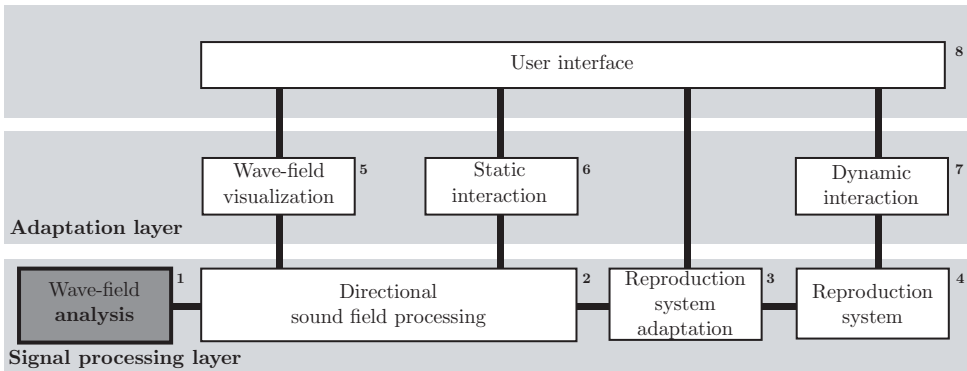


Figure 3.1: Basic block diagram of the system developed in this thesis. The part discussed in this chapter is emphasized.

3.1.1 Spatial Sampling

Sensor arrays sample the acoustic field spatially. Spherical microphone arrays in particular perform a spatial sampling of functions defined on a sphere. Spatial sampling requires limited bandwidth of the signal to avoid aliasing. Several sampling methods exist which offer a trade-off between the required number of microphones and their arrangement. Following is an overview of possible sphere-related configurations.

Rigid sphere configuration The rigid sphere is often a preferred array configuration due to its simplicity in terms of microphone mounting, as well as numerical robustness (Meyer and Elko, 2002). For broadband applications and large extrapolation areas, however, large radii are required. This makes the practical handling of such a system difficult. For this reason rigid spheres are not used in this work. A detailed discussion on rigid sphere arrays can be found in (Teutsch, 2007), (Rafaely, 2004), (Rafaely, 2005a), (Rafaely, 2005b), and (Park and Rafaely, 2005).

Hemispherical configuration Li and Duraiswami (2005) presented a hemispherical configuration of microphones, which required mounting on a rigid surface. This configuration takes advantage of the symmetries of the reflected field. Room acoustic measurements in concert halls are one of the main applications in their work. A proper placement of such a configuration cannot be

guaranteed in the desired application of this work. For auralization purposes the wave field should ideally be captured at the ear-plane of the listeners. To use the hemispherical configuration, a placement on a large surface (e.g., the floor) is required. This is difficult to realize in a concert hall.

Open sphere dual radius configuration An open sphere configuration can be useful from a practical point of view. Especially if this configuration is realized as a virtual array using a single microphone traveling along a number of positions. In this case the microphone is mounted on a computer-controlled positioning device as is explained in detail in Section 3.1.2. A single open sphere configuration using pressure measurements suffers from numerical problems, which will be presented and analyzed in the subsequent sections. Balmages and Rafaely (2007) proposed a processing method based on a dual sphere design. In terms of the desired broadband application of this project, a high number of positions is required, which will be doubled using their approach to solve numerical problems. For this reason this approach is not applicable in this work, even if it has been used successfully in room acoustic applications (e.g., (Rafaely et al., 2007a)).

Open sphere cardioid configuration The use of an open sphere with cardioid microphones provides a solution to the numerical problem of the zeros in the spherical Bessel function without sampling on two different radii (Teutsch, 2007) (Balmages and Rafaely, 2007). This will be discussed in detail in Section 3.2. As a result the complexity, as compared to the open sphere dual radius configuration, is reduced by a factor of two. This is important since a very high number of sampling points is required for a broadband application. Furthermore, an open sphere configuration can be realized using a large radius. This results in a sufficient spatial resolution over a wide frequency range. For these reasons *this configuration is used in this work*.

Spherical shell configuration To overcome the high number of microphone positions required for the previous approach, a configuration using a spherical shell has been proposed by Rafaely (2008b). For the application within this work the distribution on a shell is difficult to achieve using a virtual array.

Free sampling configuration To extend the spherical shell configuration, the next logical step is a free sampling by placing microphones at arbitrary positions in three-dimensional space. Such a configuration can be optimized by numerical methods under certain boundary conditions. The reader is referred to the publications of Li and Duraiswami (2007), Laborie et al. (2003), and Moreau et al. (2006), among others, for this approach.

3.1.2 Virtual Sensor Arrays

The summary of spherical sampling methods has shown the open sphere cardioid configuration to be most applicable in the context of this work. One of the most acoustically transparent solutions for an open-sphere configuration is the use of a virtual array. In such a system a single microphone is mounted on a robot arm and the positions on the sphere are measured subsequently. It can be extended to measure several grid points simultaneously, if a symmetrical grid and more microphones are used. Furthermore, different radii can be measured within a single measurement pass. Figure 3.2 presents the concept of the virtual sensor array. The application of a virtual array can lead to positioning errors. The cardioid microphones used have a frequency dependent directivity, as well as a limited signal-to-noise ratio (SNR). These important issues are investigated in order to identify the acceptable error range for applications desired in this work.

3.1.3 Beamforming

After the acoustic field is measured using an array configuration as discussed in the previous section, appropriate signal processing is required. It should analyze the acoustic field in a direction-dependent way. This can also be interpreted as *spatial filtering* or beamforming. (i.e., discriminating between signals based on the physical location of their sources (Veen and Buckley, 1988)). Different approaches can be identified. For a detailed discussion of the several approaches the reader is referred to Brandstein and Ward (2001) and van Trees (2002) among others.

Regular beam pattern A regular beam pattern is symmetrical around the steering direction. In case of an infinite resolution, a perfect decomposition of

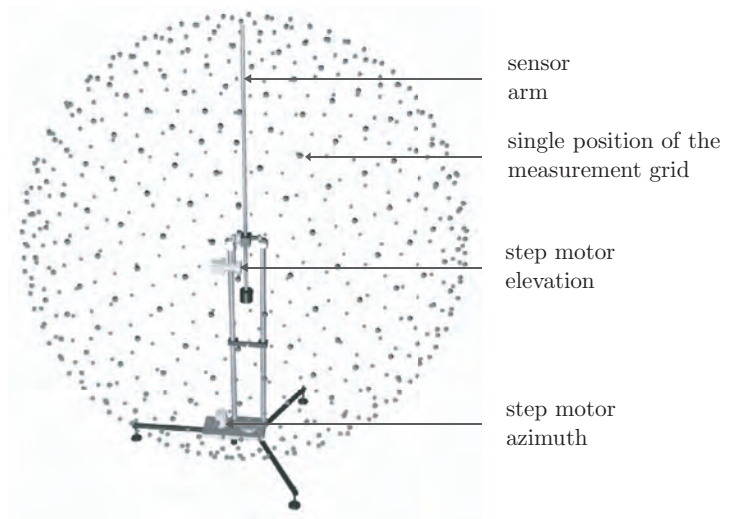


Figure 3.2: Three-dimensional visualization of the virtual sensor array used in this thesis including a set of measurement positions (Schlesinger, 2006).

the sound field into plane wave components is achieved. *This approach is used in this work.*

Delay-and-sum beam pattern A very common beamforming technique applied to different geometries of arrays is the delay-and-sum approach. An attractive property is the constant white-noise-gain. Delay-and-sum beamforming can be achieved by applying weights to the sensors that correspond to the free-field travel time differences of a plane wave in the desired direction (Rafaely, 2005b).

Dolph-Chebyshev beam pattern This beam pattern achieves the lowest side-lobe level for a given array order and main lobe width (van Trees, 2002).

Near-field beam pattern The beam patterns described above are designed under the assumption that all sources are in the far-field, such that the waves are assumed to be plane waves. In case of sources close to the array the plane wave assumption does not hold. Studies of near field sources and adequate array processing can be found in (Fisher and Rafaely, 2008). Meyer and Elko (2006) have presented an array design for near field sources.

3.1.4 Outline of the Chapter

This work studies open-sphere microphone arrays. Because of the interaction process described in Chapter 6, a plane wave decomposition of the sound field is desired. The basis for the plane wave decomposition is given by the spherical harmonic decomposition and discussed in Section 3.2. The acoustic field is always sampled in a limited number of points. The implication of this will be analyzed in Section 3.3. After the plane wave decomposition is given in Section 3.4, the measurement errors and the consequences for the desired application are analyzed in Section 3.5. The extension of bandwidth of an array measurement using multiple radii is studied in Section 3.6. In addition to the theoretical analysis an empirical analysis based on simulations and measurements in an anechoic chamber is presented in Section 3.7. The analysis of practical measurement errors is carried out with focus on their influence on the plane wave decomposition.

3.2 Spherical Harmonic Decomposition

For the application of spatial sound design it is required that a measured acoustic field is decomposed into its plane wave components. In contrast to other decompositions of a field, the plane wave decomposition is intuitively easy to understand for potential users of such a system. The superposition of plane waves form an acoustic field in a certain analyzed point and can be easily visualized. This way one can get an understanding what the acoustic result of a superposition might be. Another method of decomposition of an acoustic field is the spherical harmonic decomposition. In this case the field is decomposed into spherical harmonics of different order which allow the reconstruction of a field by superposition. A disadvantage is that the results cannot be intuitively understood by a user. Nevertheless, the spherical harmonic decomposition can provide the basis for a plane wave decomposition. Furthermore, some desired operations which need not to be visualized can be calculated directly in the spherical harmonic domain. The spherical harmonic decomposition will be derived below. In case of *interior* problems (Figure 2.3(a) illustrates the geometry), a suitable solution of the wave equation given by Eq. (2.35) must

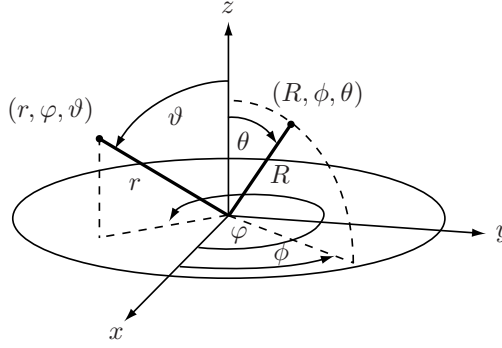


Figure 3.3: Geometry used in this chapter.

be finite in the origin. The solution was derived in Eq. (2.38):

$$P(r, \varphi, \vartheta, k) = P(\mathbf{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kr) Y_n^m(\varphi, \vartheta), \quad (3.1)$$

where $P(r, \varphi, \vartheta, k)$ denotes the sound pressure in point \mathbf{x} , k is the wavenumber, $A_{nm}(k)$ are the coefficients of the spherical harmonics $Y_n^m(\varphi, \vartheta)$ of level n and mode m , and $j_n(kr)$ is the spherical Bessel function of the first kind and the same level n . Figure 3.3 presents the geometry used in this chapter. The pressure field $P(\mathbf{x}, k)$ can be described completely using the coefficients $A_{nm}(k)$. The coefficients can be obtained using the orthonormality of the spherical harmonics (compare section A.3.3). For the radial velocity component $V_r(r, \varphi, \vartheta, k)$ a similar expression can be obtained using Eq. (3.1) and Euler's equation:

$$\mathbf{V}(\mathbf{x}, k) = \frac{1}{ick\rho_0} \nabla P(\mathbf{x}, k), \quad (3.2)$$

resulting in

$$V_r(\mathbf{x}, k) = \frac{1}{ic\rho_0} \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n'(kr) Y_n^m(\varphi, \vartheta). \quad (3.3)$$

In this equation $j_n'(kr)$ is the derivative of the spherical Bessel function with respect to kr given by

$$j_n'(kr)k = \frac{\partial j_n(kr)}{\partial r} = \frac{\partial j_n(kr)}{\partial kr} \frac{\partial kr}{\partial r}. \quad (3.4)$$

Multiplication of Eq. (3.1) with the conjugate spherical harmonic function $\overline{Y_n^m(\varphi, \vartheta)}$ and integration over the surface of the unit sphere leads to (Williams, 1999, p. 218)

$$A_{nm}(k) = \frac{1}{j_n(kr)} \int_0^{2\pi} \int_0^\pi P(\mathbf{x}, k) \overline{Y_n^m(\varphi, \vartheta)} \sin \vartheta d\vartheta d\varphi. \quad (3.5)$$

The integral in Eq.(3.5) corresponds to the spherical wave spectrum defined as

$$P_{nm}(r, k) = \int_0^{2\pi} \int_0^\pi P(\mathbf{x}, k) \overline{Y_n^m(\vartheta, \varphi)} \sin \vartheta d\vartheta d\varphi. \quad (3.6)$$

It is important to note that computing the spherical wave spectrum $P_{nm}(r, k)$ is equivalent to a spherical Fourier transform of $P(r, \varphi, \vartheta, k)$ as given in Eq. (A.22). Based on these equations, a measured field can be decomposed into orthogonal eigenfunctions. For the development of plane wave decomposition it is advantageous to study the expansion of plane waves into spherical harmonics.

Plane wave expansion The pressure field of an ideal plane wave at the position $\mathbf{x} = (r, \varphi, \vartheta)$ in spherical coordinates as given in Appendix A.1 reads (Ziomek, 1995)

$$P(\mathbf{x}, k) = P_0(k) e^{-i\mathbf{k} \cdot \mathbf{x}}. \quad (3.7)$$

With

$$\mathbf{k} \cdot \mathbf{x} = kr [\sin \vartheta \sin \vartheta_0 \cos(\varphi - \varphi_0) + \cos \vartheta \cos \vartheta_0], \quad (3.8)$$

this results in

$$P(r, \varphi, \vartheta, k) = P_0(k) e^{-ikr[\sin \vartheta \sin \vartheta_0 \cos(\varphi - \varphi_0) + \cos \vartheta \cos \vartheta_0]} \quad (3.9)$$

where $P_0(k) = 1$ is the spectrum of the plane wave, and \mathbf{k} is the wave vector in spherical coordinates. The wave vector represents the direction of incidence of the plane wave specified by (φ_0, ϑ_0) . Combining Eq. (3.7) and Eq. (3.1) yields the expansion of a plane wave in terms of spherical harmonics for an open sphere array utilizing omni-directional sensors

$$e^{-i\mathbf{k} \cdot \mathbf{x}} = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kr) Y_n^m(\vartheta, \varphi). \quad (3.10)$$

Following (Williams, 1999, p. 227) this expression can be transformed into

$$e^{-i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{m=-n}^n Y_n^m(\varphi, \vartheta) \overline{Y_n^m(\varphi_0, \vartheta_0)}. \quad (3.11)$$

Comparing Eq. (3.11) and Eq. (3.10), the coefficients A_{nm} can be identified as

$$A_{nm}(k) = 4\pi i^n \overline{Y_n^m(\varphi_0, \vartheta_0)}. \quad (3.12)$$

If there is no interaction between the sound field and the aperture used for the measurement, Eq. (3.12) leads to correct results. Using Eq. (3.12) the ideal spherical harmonic coefficient up to order n for a single plane wave under free-field conditions can be calculated. This is important for the evaluation of measurement errors in order to compare non-ideal scenarios with the ideal situation. The problems of measurement errors are discussed in Section 3.5.

Spherical array measurements The spherical harmonics coefficients $A_{nm}(k)$ can be determined if the sound pressure $S(\mathbf{x}, k) = S(r, \varphi, \vartheta, k)$ on the surface of the sphere is known. To generalize Eq. (3.3) and Eq. (3.1) to different array designs as described in Section 3.1.2, the following extension is used:

$$S(\mathbf{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) b_n(kr) Y_n^m(\vartheta, \varphi), \quad (3.13)$$

where $S(\mathbf{x}, k)$ is the response of the sensor element of the array. The coefficients are calculated by integrating over the sphere, resulting in

$$A_{nm}(k) = \frac{1}{b_n(kr)} \int_0^{2\pi} \int_0^{\pi} S(\mathbf{x}, k) \overline{Y_n^m(\vartheta, \varphi)} \sin \vartheta d\vartheta d\varphi. \quad (3.14)$$

In case of open sphere arrays with omnidirectional sensors the sensor response is equivalent to the sound pressure $P(r, \vartheta, \varphi, k)$. The new expression $b_n(kr)$ is defined as mode strength and depends on the array design used. It is given for an omni open sphere array by Balmages and Rafaely (2007) (Teutsch, 2007):

$$b_n(kr) = j_n(kr), \quad (3.15)$$

The magnitude of the mode strength $b_n(kr)$ is plotted in Figure 3.4 for several levels n . It can be noticed that for small kr only the zero-level is present. For

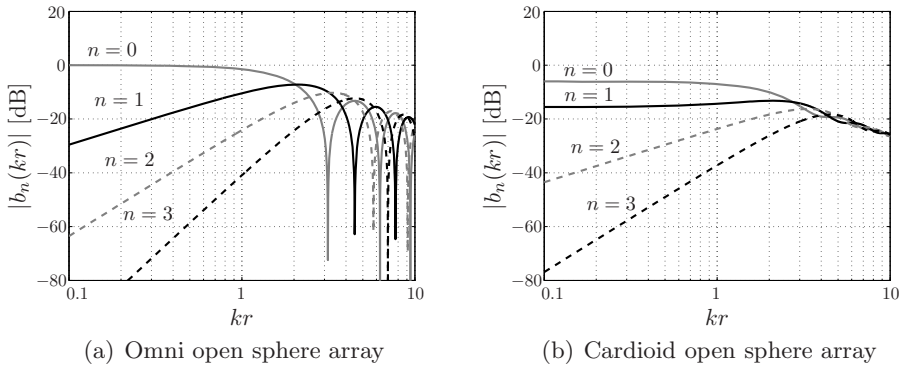


Figure 3.4: Mode strength $b_n(kr)$ for spherical microphone arrays with different designs and levels $n = 1$ to $n = 3$ (Thiergart, 2007).

higher values of kr more levels arise. It is shown that all levels $n < \lceil kr \rceil$ should have sufficient strength. The strength for levels $n > \lceil kr \rceil$ decays rapidly for increasing n . It is clear from Eq. (3.14) that a low mode strength $b_n(kr)$ involves a high amplification of the integral expression, which can lead to numerical problems. For this reason the maximal used level, N_{max} , is defined as

$$N_{max} = \lceil kr \rceil + \iota, \quad (3.16)$$

where $\iota \in [0, \infty]$. The choice of ι depends on the desired robustness of the calculations. This point will be discussed in detail in the following sections. The mode strength of an omni open sphere array is presented in Figure 3.4(a). For higher values of kr several singularities arise due to the zeros of the spherical Bessel function. The result can be excessive noise amplification for several frequencies as shown in Section 3.5. Two solutions are proposed in the literature to overcome the numerical problems of an open-sphere sensor array (Balmages and Rafaely, 2007). These are the use of two slightly different radii and choosing the radii without zeros for a specific frequency or the use of directional sensor apertures. This thesis studies the possibility to use directional sensor apertures. The signal acquired by a first-order sensor can be given by a combination of pressure $P(\mathbf{x}, k)$ and velocity in the normal direction of the sensor $V_r(\mathbf{x}, k)$ (i.e., in radial direction in case of a spherical sensor aperture), resulting in (Poletti, 2005):

$$S(\mathbf{x}, k) = \beta P(\mathbf{x}, k) + (1 - \beta) [\rho_o c V_r(\mathbf{x}, k)], \quad (3.17)$$

where $\beta \in [0, 1]$ is a parameter to fade from omni-directional to figure-of-eight characteristic equivalent to a pure pressure measurement and a velocity measurement in radial direction. For $\beta = 0.5$ Eq. (3.17) yields a cardioid microphone characteristic. For a cardioid sensor array the following mode strength $b_n(kr)$ has to be used (Poletti, 2005):

$$b_n(kr) = \frac{1}{2} [j_n(kr) - ij'_n(kr)]. \quad (3.18)$$

The definition of the mode strength can now be extended to

$$b_n(kr) = \begin{cases} j_n(kr) & \text{omni open sphere} \\ \frac{1}{2}(j_n(kr) - ij'_n(kr)) & \text{cardioid open sphere.} \end{cases} \quad (3.19)$$

Figure 3.4(b) shows the mode strength $b_n(kr)$ for the cardioid open sphere aperture. No singularities are present because the spherical Bessel function and its derivative have no common zeros. As a result the use of first order sensors is a suitable way to overcome the numerical problems of omni open sphere array designs. It should also be noted that the mode strength for level $n = 1$ is boosted in comparison to the open-sphere design. The first level is a natural characteristic of the cardioid microphone (Teutsch, 2007). Practical problems and limitations of the use of cardioid microphones are studied in detail in the following sections.

3.3 Discrete Spherical Apertures

In practical applications the sound field can only be measured at a finite number of microphone positions. As a result the determination of the coefficients $A_{nm}(k)$ requires an approximation of the integral by a finite summation. The approximation of a continuous integrals is known as quadrature. An integral of a function $f(\phi, \theta)$ over the surface of the unit sphere S^2 can be approximated by

$$\int_{S^2} f(\phi, \theta) d\Omega = \int_0^{2\pi} \int_0^\pi f(\phi, \theta) \sin \theta d\theta d\phi \approx \sum_{q=1}^Q f(\phi_q, \theta_q) w_q \quad (3.20)$$

Q denotes the number of sampling points and w_q are the quadrature weights depending on the sampling scheme. As a result the integral in Eq. (3.14) is

approximated by a finite sum

$$A_{nm}(k) \approx \hat{A}_{nm}(k) = \frac{1}{b_n(kr)} \sum_{q=1}^Q w_q S_q(\mathbf{x}_q, k) \overline{Y_n^m(\varphi_q, \vartheta_q)}. \quad (3.21)$$

where $\hat{\cdot}$ indicates the approximation using a finite sum. Like sampling in the time domain, sampling on a sphere requires a limited bandwidth corresponding to a limited number of coefficients $A_{nm}(k)$ stating that

$$A_{nm}(k) = 0 \quad \forall n > N_{max}, \quad (3.22)$$

where N_{max} denotes the maximum level of the coefficients of the sound field. Eq. 3.22 must be ensured before sampling, otherwise spatial aliasing will corrupt the coefficients in a way that a subsequent filtering cannot compensate for. A detailed discussion on spatial aliasing in spherical microphone arrays can be found in (Rafaely et al., 2007b). For the sampling of level-limited sound fields defined by (3.22) several quadratures are available. Following (Galdo, 2007) three commonly used sampling schemes as presented in Figure 3.5 are:

1. The *Chebyshev* quadrature is characterized by a uniform sampling in co-elevation ϑ and azimuth φ .
2. The *Gauss-Legendre* quadrature results in a uniform sampling in azimuth direction φ . In co-elevation ϑ the sampling uses the zeros of the Legendre polynomials of the level $N_{max} + 1$.
3. The *Lebedev* quadrature is nearly uniformly spaced over the sphere surface.

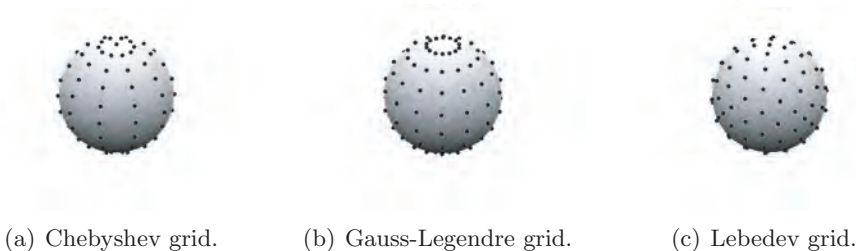


Figure 3.5: Sampling schemes for aliasing-free sampling of level-limited fields on a spherical aperture ($Q = 110$ for (a), (c) and $Q = 112$ for (b)).

Chebyshev Quadrature The Chebyshev quadrature requires a uniform sampling in co-elevation and azimuth. Due to the simplicity of the rule, it is used frequently in measurement applications. The grid points of a Chebyshev grid of order $M = 2N + 1$ are defined by $\mathcal{G}_{\text{Ch}}^{(M)}$ as

$$\mathcal{G}_{\text{Ch}}^{(M)} = \left\{ (\varphi_u, \vartheta_v) : \varphi_u = \frac{\pi u}{M-1}; \vartheta_v = \frac{2\pi v}{M-1}; \right. \\ \left. u, v \in \mathbb{Z}; u \in [0, M-1]; v \in \left[-\frac{M-1}{2}, \frac{M-1}{2} \right) \right\}, \quad (3.23)$$

where M can only be odd. The number of grid points to compute the spherical modes between $-\frac{M-1}{2}$ and $\frac{M-1}{2}$ is $Q_{\text{Ch}} = M \times (M-1)$. The quadrature weights are calculated as (Galdo, 2007)

$$w_{u, \text{Ch}}^{(M)} = \epsilon_u^{M-1} \frac{4\pi}{M-1} \sum_{s=0}^{(M-1)/2} \epsilon_u^{(M-1)/2} \frac{2}{1-4s^2} \cos\left(\frac{us2\pi}{M-1}\right), \quad (3.24)$$

$$w_{v, \text{Ch}}^{(M)} = \frac{1}{M-1}, \quad (3.25)$$

where

$$\epsilon_u^{(M)} = \begin{cases} 0.5 & \text{for } u = 0, M \\ 1 & \text{for } u = 1, 2, \dots, M-1 \end{cases}. \quad (3.26)$$

Since the Chebyshev quadrature points can be organized in a regular lattice, the sum in Eq. (3.20) can be rewritten as two sums

$$\int_{S^2} f(\phi, \theta) d\Omega \approx \sum_{u=0}^M w_{u, \text{Ch}}^{(M)} \sum_{v=-(M-1)/2}^{M-3/2} w_{v, \text{Ch}}^{(M)} \cdot f(\phi_u, \theta_v) \quad (3.27)$$

An advantage of the Chebyshev grid is the regular angle difference between the quadrature points. This can be useful when spatial samples are taken by a virtual microphone array. The disadvantage is the large number of samples compared to the other two sampling schemes.

Gauss-Legendre Quadrature The Gauss-Legendre quadrature is sampled at the zeros of the Legendre polynomials in co-elevation and uniformly in az-

imuth. Its grid points of order M are defined by $\mathcal{G}_{GL}^{(M)}$ as

$$\mathcal{G}_{GL}^{(M)} = \left\{ (\varphi_u, \vartheta_v) : \varphi_u = \frac{\pi u}{M-1}; \vartheta_v = \arccos \left(z_u^{(\frac{M-1}{2})} \right); \right. \\ \left. u, v \in \mathbb{Z}; u \in \left[0, \frac{M-1}{2} \right]; v \in \left[-\frac{M-1}{2}, \frac{M-1}{2} \right] \right\}, \quad (3.28)$$

where z_u^i are the zeros of the Legendre polynomials of order $i+1$. The quadrature weights $w_{u,GL}^{(M)}$ along the longitudes¹ can be computed with the Gaussian quadrature rule (Engeln-Müllges et al., 2005). The weights along the latitudes² $w_{v,GL}^{(M)}$ are the same as for the Chebyshev quadrature. The total number of samples for the Gauss-Legendre grid is $N_{GL} = \frac{M+1}{2} \times (M-1)$. The quadrature problem can be solved using

$$\int_{S^2} f(\phi, \theta) d\Omega \approx \sum_{u=0}^M w_{u,GL}^{(M)} \sum_{v=-(M-1)/2}^{M-3/2} w_{v,GL}^{(M)} f(\phi_u, \theta_v). \quad (3.29)$$

Lebedev Quadrature An intuitive grid is a uniform sampling on the sphere. This means that each sampling point has the same distance to its nearest neighbors. As a consequence the points will not lie on a lattice separable in azimuth and co-elevation. The Lebedev grid given in (Lebedev, 1975, 1976, 1977) and (Lebedev and Skorokhodov, 1992; Lebedev and Laikov, 1999; Lebedev, 1995) approaches an ideal uniform distribution of samples. Finding the samples is complicated and no general formula can be given. Grid points and weights can be calculated using a Fortran code given in (Lebedev, 2009) up to order $M = 131$. For a given order, the number of grid points can be approximated with $N_{Lb} \approx \frac{(M+1)^2}{3}$. For this grid the quadrature problem can be solved using

$$\int_{S^2} f(\phi, \theta) d\Omega = \int_0^{2\pi} \int_0^\pi f(\phi, \theta) \sin \theta d\theta d\phi \approx \sum_{q=0}^{N_{Lb}-1} f(\phi_q, \theta_q) w_q \quad (3.30)$$

where $\{\phi_q, \theta_q\} \in \mathcal{G}_{Lb}^{(M)}$. The main advantage of this quadrature is the reduced number of microphone positions compared to the Chebyshev and Gauss-

¹co-elevation

²azimuth

Legendre quadrature.

Comparison of the Quadratures The quadratures introduced are used to integrate in case of a spherical harmonic transform for a grid of order M . Grids can be applied for spherical harmonics with level $n \leq N$, where $N = 2M$. The relationship between the number of sampling points for different quadratures for any given order M can be approximated as (Galdo, 2007)

$$N_{Ch} \approx 2N_{GL} \approx 3N_{Lb}. \quad (3.31)$$

The *Chebyshev* grid is the simplest grid for measurements since sampling both angles uniformly allows accurate positioning of the sensor in the case of virtual arrays. It is the most inefficient quadrature grid with respect to the number of samples required. As a result, the measurement process will be much more time-consuming when using a Chebyshev grid for the same resolution, as compared to other grids. The advantage of the *Gauss-Legendre* grid is that it is the optimal solution for the co-elevation. The angles are separable in both directions which makes the calculation efficient using matrix calculus. However, this grid requires more sampling points on the sphere than the Lebedev grid. The *Lebedev* quadrature cannot be separated into weights along the azimuth and co-elevation direction because of its true bivariant nature. The major advantage is the reduced number of grid points required. This is especially important in the context of this work as high orders are required because of the desired broadband measurement and analysis. However, processing both directions lead to a higher computational complexity. In terms of room acoustic analysis the main focus is the reduction of measurement time and for this reason *the Lebedev quadrature is the solution used.*

Spatial Aliasing The quadratures require a level-limited sound field in order to provide aliasing-free sampling. However, even simple sound fields such as single plane wave fields are not restricted to a certain maximum level N_{\max} . For this reason spatial aliasing must be expected. This follows directly from Eq. (3.10). A plane wave involves an infinite number of non-zero coefficients $A_{nm}(k)$. Nevertheless, as stated in Appendix A.3.2, the spherical Bessel functions of the first kind $j_n(kr)$ decay rapidly for $kr > n$. Hence, the strength of the coefficients in Eq. (3.12) decays rapidly for $kr > n$. As a result, the aliasing error can be neglected if the operating frequency of the microphone array fulfills $kr \ll N_{\max}$. This restriction shows that spatial aliasing limits the

array performance at high frequencies or for large radii, respectively. It will be shown in the subsequent sections that the following condition is sufficient to avoid aliasing:

$$[kr] \leq N_{\max}. \quad (3.32)$$

To prove the quadratures, an expression to calculate the spatial aliasing error is derived. Substituting Eq. (3.14) into Eq. (3.21) and re-arranging the terms yields

$$\hat{A}_{nm}(k) = \frac{1}{b_n(kr)} \cdot \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} A_{n'm'}(k) \cdot b_{n'}(kr) \cdot \left[\sum_q^Q w_q \cdot Y_{n'}^{m'}(\Omega_q) \cdot \overline{Y_n^m(\Omega_q)} \right]. \quad (3.33)$$

The term in the square brackets is an approximation of the orthonormality condition of the spherical harmonics in Eq. (A.19). Hence,

$$\sum_q^Q w_q \cdot Y_{n'}^{m'}(\Omega_q) \cdot \overline{Y_n^m(\Omega_q)} = \delta_{n',n} \delta_{m',m} + \epsilon_a(n, m, n', m'), \quad (3.34)$$

where $\delta_{n',n}$ and $\delta_{m',m}$ are Kronecker deltas (Rafaely et al., 2007b). If $\epsilon_a(n, m, n', m')$ is zero, the exact integral is obtained. The function $\epsilon_a(n, m, n', m')$ denotes the aliasing error. Equation (3.34) is computed for a Lebedev grid with $Q = 110$

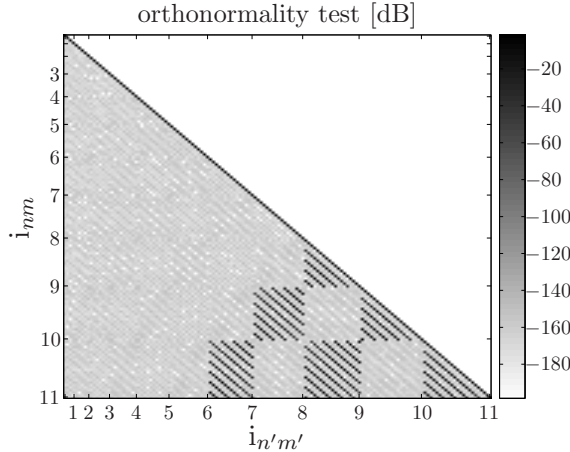


Figure 3.6: Orthonormality test for spherical harmonics up to level $n = 11$ for $Q = 110$ points on a Lebedev grid. Significant aliasing arises for $n > 8$ (Thiergart, 2007).

points for all combinations of n', n and m', m up to level $n = 11$. The result is depicted in Figure 3.6 and is referred as *orthonormality test*. The index i_{nm} specifies the levels of the coefficients $\hat{A}_{nm}(k)$, e. g., between $i_{nm} = 9$ and $i_{nm} = 10$ all modes m of level $n = 10$ are present. It is expected from (3.34) that if no aliasing occurs, the orthogonality test must be 1 (0 dB) for $i_{nm} = i_{n'm'}$ and 0 ($-\infty$ dB) otherwise. This constraint is fulfilled for $n \leq 8$. For higher levels, significant aliasing arises, which confirms Eq. (3.32). A different way to reduce spatial aliasing other than denser sampling is to apply a spatial low-pass filter. Referring to (Meyer and Agnello, 2003) and (Rafaely et al., 2007b), this filtering is equivalent to a spatial integration of the sound pressure. As stated above, the low-pass filter must be employed before sampling the sphere. This can be realized by using microphones with large membranes covering a wide area of the sphere. Conventional microphones are not suitable for this task (Meyer and Agnello, 2003), (Rafaely et al., 2007b). For this reason, and since the microphone array used in this work can provide a high number of microphone positions, a denser sampling is applied to reduce spatial aliasing.

3.4 Plane Wave Decomposition

After the spherical harmonic decomposition and the effect of sampling have been studied, this representation is now transformed into a plane wave decomposition. A possible way has been developed by Rafaely (2004). The approach is based on spherical Fourier transform (SFT) as given in Appendix A.3.3. The coefficients of the SFT $f_{nm}(k)$ and the coefficients of the spherical harmonics $A_{nm}(k)$ are related by

$$f_{nm}(k) = A_{nm}(k)b_n(kr), \quad (3.35)$$

where $b_n(kr)$ is the mode strength. First a sound field consisting of a single plane wave with a direction of arrival (DOA) (φ_0, ϑ_0) is considered. The corresponding spherical harmonic coefficients $A_{nm}(k)$ can be obtained from Eq. (3.12). Inserting the result in Eq. (3.35) leads to

$$f_{nm}(k) = 4\pi i^n b_n(kr) \overline{Y_n^m(\varphi_0, \vartheta_0)}. \quad (3.36)$$

Now an infinite number of plane waves with the magnitude spectrum $w(\varphi_0, \vartheta_0, k)$ are considered. The spherical Fourier coefficients of the sound field can be cal-

culated by integrating over all directions of incidence

$$f_{nm}(k) = 4\pi i^n b_n(kr) \int_0^{2\pi} \int_0^\pi w(\varphi_0, \vartheta_0, k) \overline{Y_n^m(\varphi_0, \vartheta_0)} \sin \vartheta d\vartheta d\varphi. \quad (3.37)$$

The integral is the SFT of the amplitude weights $w(\varphi_0, \vartheta_0)$ and yields the Fourier coefficients $w_{nm}(k)$

$$w_{nm}(k) = f_{nm}(k) \frac{1}{4\pi i^n b_n(kr)}. \quad (3.38)$$

The amplitude $w_s(\varphi_s, \vartheta_s, k)$ of a plane wave arriving from a specific direction (φ_s, ϑ_s) can be determined by performing an inverse spherical Fourier transform (ISFT) of Eq. (3.38)

$$w_s(\varphi_s, \vartheta_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm}(k) \frac{1}{4\pi i^n b_n(kr)} Y_n^m(\varphi_s, \vartheta_s). \quad (3.39)$$

The function $w_s(\varphi_s, \vartheta_s, k)$ is the *directivity function* (or direction weights) representing the result of the plane wave decomposition (PWD). (φ_s, ϑ_s) defines the steering direction of the aperture, denoting the direction for which the PWD is computed. Substituting Eq. (3.35) into Eq. (3.39) gives the final expression (Rafaely, 2004):

$$w_s(\varphi_s, \vartheta_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{1}{4\pi i^n} A_{nm}(k) Y_n^m(\varphi_s, \vartheta_s), \quad (3.40)$$

which gives the PWD based on the spherical harmonic coefficients $A_{nm}(k)$. The direction weights $w_s(\varphi_s, \vartheta_s, k)$ can be interpreted as an impulse response in the direction (φ_s, ϑ_s) , which is an intuitive understanding of the plane wave decomposition. The superposition of all impulses responses, corresponding to all DOA, construct the sound field in the analyzed point. The term *plane wave decomposition* is somewhat misleading because it supposes an infinite resolution. This is only correct from a theoretical point of view. In practice the order of analysis is always limited to a certain value N_{max} . This will lead to a limited spatial resolution. The spatial resolution of a plane wave decomposition is studied in the next section.

3.4.1 Spatial Resolution

The resolution of the plane wave decomposition (PWD) is limited, as the peak in the directional weights $w_s(\varphi_s, \vartheta_s, k)$ resulting from a single plane wave is not infinitesimally thin. The spatial width of the peak is determined only by the maximum level N of the coefficients $A_{nm}(k)$ within the calculations, at least in theory. However, in this section it is demonstrated that the PWD resolution is reduced by measurement errors. To illustrate this, a sound field of a single plane wave with frequency $f = 1574 \text{ Hz}$ and arrival direction $\Omega_0 = (\varphi_0, \vartheta_0) = (0^\circ, 90^\circ)$ is considered. The sound field is simulated using a cardioid open-sphere array with a Lebedev quadrature using $Q = 170$ microphone positions and a radius $r = 0.3245 \text{ m}$. The PWD is computed for $N = 10$ and $N = 5$. The results depicted in Figure 3.7 show that the peaks become wider for lower maximum levels N . Therefore, the resolution of the PWD decreases for decreasing N . The following considerations do not de-

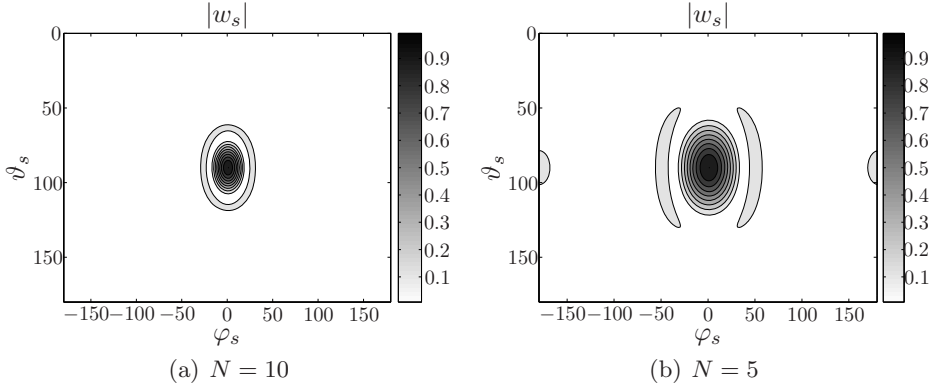


Figure 3.7: Direction weights w_s for a sound field of a single plane wave, computed for different maximum levels N . The absolute values are plotted. The results are normalized to one (Thiergart, 2007).

pend on a particular array design. The spherical harmonic coefficients $A_{nm}(k)$ can be computed directly for a single plane wave sound field using Eq.(3.12). The corresponding directional weights w_s are found from Eq.(3.40). Inserting Eq.(3.12) into Eq.(3.40) yields

$$w_s(\varphi_s, \vartheta_s) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \overline{Y_n^m(\varphi_0, \vartheta_0)} Y_n^m(\varphi_s, \vartheta_s), \quad (3.41)$$

where (φ_0, ϑ_0) is the direction of arrival (DOA) of the plane wave and (φ_s, ϑ_s) denotes the steering direction of the microphone array. According to the spherical harmonic addition theorem as given in Eq.(A.24), Eq.(3.41) can be rewritten as

$$w_s(\varphi_s, \vartheta_s) = w_s(\Theta) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} P_n(\cos(\Theta)), \quad (3.42)$$

where $P_n(\cos(\Theta))$ is the Legendre function of level n and Θ is the angle between the directions (φ_s, ϑ_s) and (φ_0, ϑ_0) . The angle Θ is given by (Varshalovich et al., 1988, p. 163):

$$\Theta = \cos \vartheta_s \cos \vartheta_0 + \sin \vartheta_s \sin \vartheta_0 \cos(\varphi_s - \varphi_0). \quad (3.43)$$

In view of (Gradshteyn and Ryzhik, 2007, p. 986), Eq. (3.42) can be simplified to

$$w_s(\Theta) = \frac{N+1}{4\pi(\cos \Theta - 1)} [(P_{N+1}(\cos \Theta) - P_N(\cos \Theta))], \quad (3.44)$$

where N is the maximum level of the spherical harmonics. $P_N(\cdot)$ denotes the Legendre polynomial of level N . The directional weights $w_s(\Theta)$ are the result of the PWD for a sound field consisting of a single plane wave and define the spatial resolution as a function of the maximum level N . Figure 3.8(a) shows the result of Eq. (3.44) for different levels N . It can be observed that the peak of the direction weights, which represents the arrival direction of the plane wave, becomes wider for decreasing N . The first value Θ_1 for which $w_s(\Theta) = 0$ and $\Theta > 0$ is defined as half the resolution of the PWD (Rafaely, 2004). An empirical rule to determine the zero of $w_s(\Theta)$ for a certain level N is given by (Rafaely, 2004)

$$\Theta_1 \approx \frac{180^\circ}{N}. \quad (3.45)$$

The zeros Θ_1 are illustrated in Figure 3.8(b) as a function of N , computed exactly from Eq. (3.44) and approximated with Eq. (3.45). The approximation error is less than 2° for $N \in [5, 50]$. Table 3.1 shows the resolution ($= 2\Theta_1$) of the PWD for several maximum levels N , obtained from (3.45). The resolution of the plane wave decomposition is an appropriate measure to evaluate the performance of a spherical microphone array. A frequently used method (e.g. (Meyer and Elko, 2002), (Park and Rafaely, 2005)) to visualize this performance is the *directional response*. It can be determined from Eq. (3.44) by plotting the results over all angles Θ . Thus, the directional response of a microphone array is equivalent to the directivity function, or directional weights

w_s in case of a single plane wave sound field. An example is depicted in Figure 3.9 for the maximum levels $N = 10$ and $N = 5$. The main lobe of the beam pattern denotes the steering direction of the microphone array. Furthermore, the width of this lobe indicates the resolution. It can be noticed that the resolution is higher for larger N . Moreover, several side lobes and a back lobe are visible, which can also be observed in Figure 3.7.

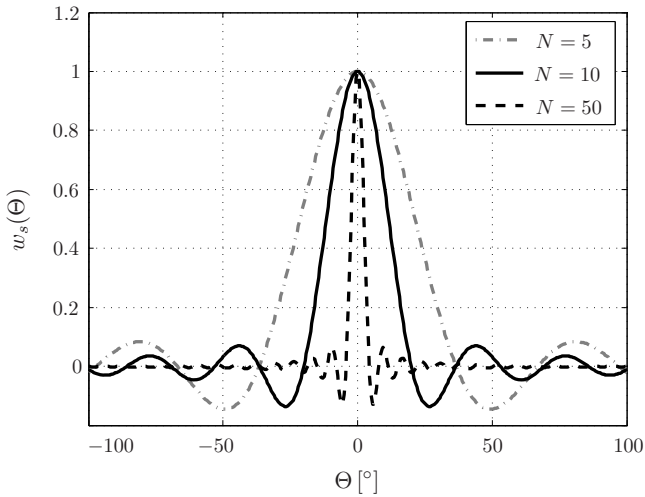
PWD Resolution [°]	Maximum level N
72	5
36	10
14.4	25
12	30
7.2	50

Table 3.1: Resolution of the PWD for different maximum levels N .

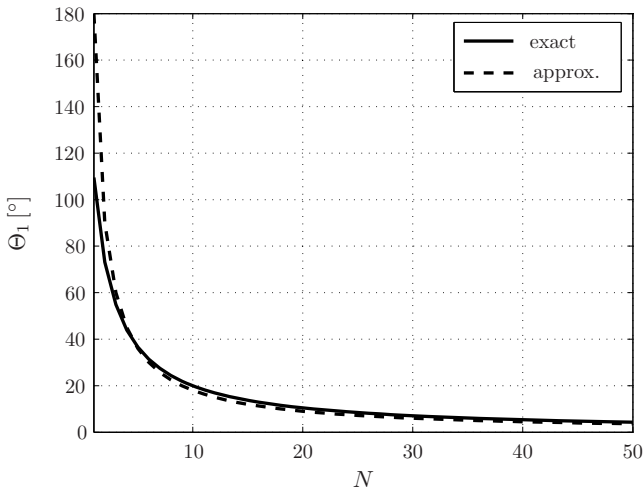
A second common measure to evaluate the performance of a microphone array is the directivity index (DIX) (Gover et al., 2002). The directivity index is the peak to average ratio of the outputs of the directivity function. It describes the variation of the energy within the output over all steering directions Ω_s . Therefore, it is a measure which can be used to characterize the directivity of an array. The directivity index is defined as (Gover et al., 2002),

$$\text{DIX} = 10 \log_{10} \left[\frac{\max(|w_s|^2)}{\text{E}(|w_s|^2)} \right], \quad (3.46)$$

where $\max(\cdot)$ denotes the maximum value and $\text{E}(\cdot)$ denotes the average value. For example, the DIX for the result in Figure 3.7(a) is approximately 21 dB, whereas for Figure 3.7(b) it is about 16 dB. The resolution of the PWD depends only on the maximum level N of the spherical harmonic coefficients, and not on the frequency f , nor on the product of wavenumber and radius kr . As a consequence, *in theory*, the same resolution can be achieved at low frequencies as at higher frequencies. However, this requires computing spherical harmonic coefficients of high levels at low values of kr (see Figure 3.4). In the next section this statement is evaluated in practical scenarios, i.e., including measurement noise on the sensor signals.



(a) Direction weights w_s



(b) Half resolution of the PWD

Figure 3.8: Plot (a) shows the (normalized) direction weights w_s as a function of Θ for a sound field of a single plane wave. Plot (b) shows the first positive zero of w_s for different maximum levels N , which equals the half resolution of the PWD (Thiergart, 2007).

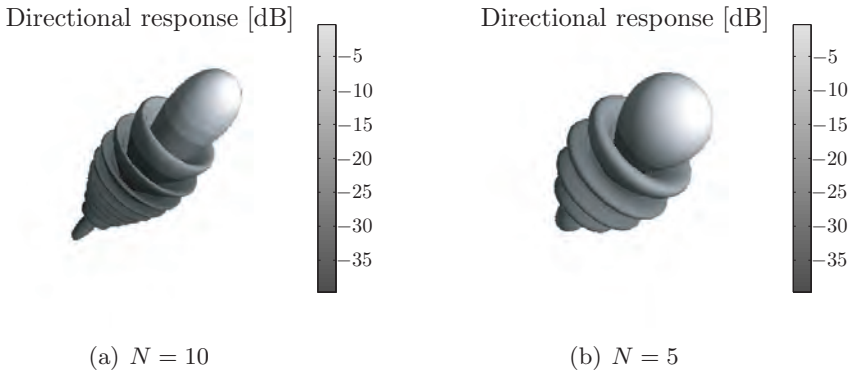


Figure 3.9: Directional response of a plane wave decomposition for different maximum levels N (Thiergart, 2007).

3.4.2 Frequency Bandwidth

The usable frequency bandwidth is an important property in view of the desired auralization application. For a high quality auralization a very large bandwidth is needed. This section analyses the frequency range in which spherical microphone arrays with a given radius r can be used. A sound field consisting of a single plane wave with a frequency range from $20 \text{ Hz} \leq f \leq 5 \text{ kHz}$ is considered. The PWD is computed for two cardioid arrays with a radius of $r = 0.3245 \text{ m}$ and $Q = 170$ sampling positions using a Lebedev grid. The first array is computed under ideal conditions and for the second, normally distributed microphone noise with a SNR of 40 dB is used. A maximum level of $N = 10$ is assumed. Figure 3.10 shows the normalized directional weights as a function of frequency for the steering directions $\Omega_s = (\varphi_s, 90^\circ)$. In the noise-free situation in Figure 3.10(a), the applicable frequency range only has an upper limit only due to spatial aliasing, which starts at around $f = 2.5 \text{ kHz}$. This is in contrast to the noisy microphone array in Figure 3.10(b), which shows a lower frequency limit. Notice that the value range of the plot is clipped to 1 for visualization reasons, but the magnitude of the directional weights w_s tends towards infinity for $f \rightarrow 0 \text{ Hz}$. The reason for the lower frequency limit is the mode strength $b_n(kr)$, which becomes small for the higher levels n at low values of kr (see Figures 3.4). The result is an excessive noise amplification which corrupts the results of the PWD. This problem is not specific to cardioid arrays. Furthermore, the results in Figure 3.10 show that Eq. (3.32) is appropriate for estimating the aliasing-free frequency limit, since $f_{max} \leq \frac{Nc}{2\pi r}$. The restricted

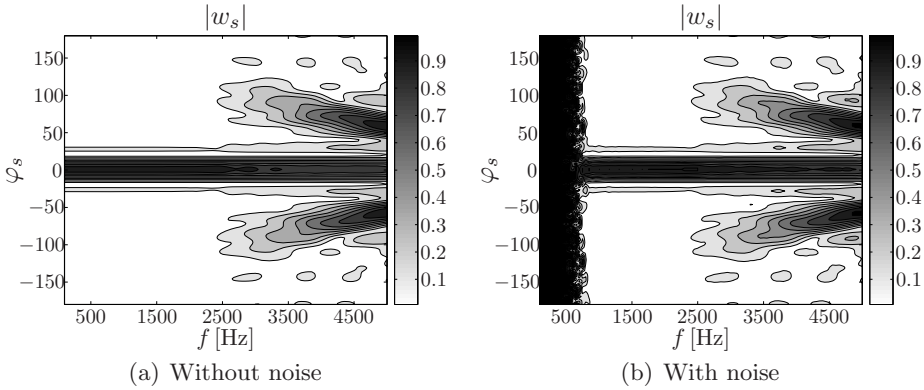


Figure 3.10: Normalized direction weights w_s for a single plane wave as a function of f . The plane wave decomposition is performed for a noise-free array and a noisy array ($SNR = 40$ dB, $r = 0.3245$ m, $N = 10$ and $Q = 170$) (Thiergart, 2007).

frequency bandwidth of microphone arrays is a major drawback for broadband auralization applications in practice. The results show that the array performance is limited at large kr due to spatial aliasing, and also at small kr due to a decreased array robustness. A solution for overcoming spatial aliasing is using a higher number of microphone positions Q . However, this can increase the measurement effort dramatically. A second way is the use of smaller array radii r , which in turn reduces the array robustness against noise at low frequencies. To improve the low frequency robustness, the maximum level N within the calculations can be lowered. This leads to a smaller resolution of the PWD. Therefore, it might be very difficult in practice to find an optimal setup which at the same time provides a high frequency bandwidth, a sufficient resolution at low frequencies, and a moderate measurement effort. These findings provide a motivation for the development of a multi-radii measurement system. The basic idea is a simultaneous measurement on different radii r , in which the larger radius is used at the lower frequencies. This method is not feasible for rigid sphere designs, which is an advantage of open sphere arrays. The combination of such measurements for a system consisting of two separate radii is discussed in Section 3.6. A second way is a varying maximum level N , which depends on the actual value of kr . In this case the resolution of the PWD varies with frequency. This is also useful in view of further measurement errors. The following section will discuss relevant types of errors which will occur in a real measurement system.

3.5 Error Analysis of Virtual Spherical Sensor Apertures

The performance of practical microphone arrays is limited by various measurement problems, which can significantly influence the results of the plane wave decomposition. Throughout this section these issues are investigated. The measurement problems can be classified following Figure 3.11. A mathematical approach is derived to evaluate the influences of sensor noise, positioning inaccuracies of the array, and spatial aliasing, on the results of the PWD. The optimal array parameters to achieve a sufficient robustness against the various measurement problems are studied. Examples of plane wave decompositions are provided to show the effects of the various errors.

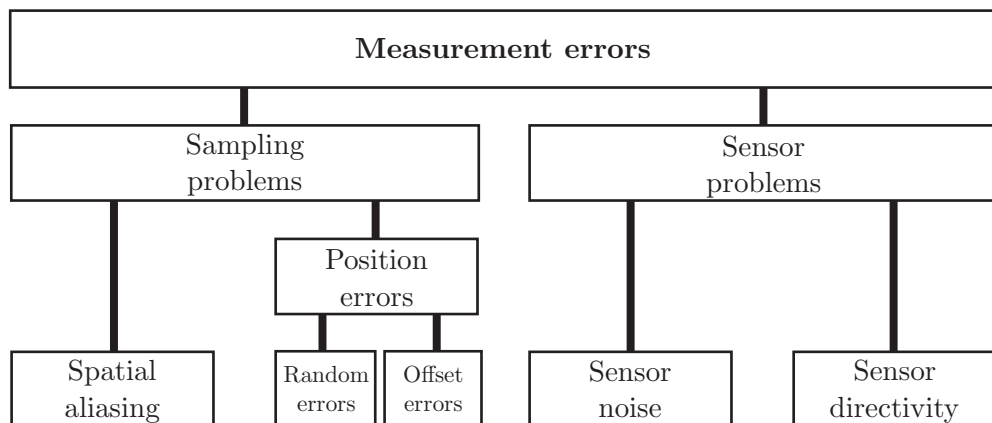


Figure 3.11: Classification of measurement errors of virtual spherical sensor apertures.

3.5.1 Description of Spatial Aliasing and Measurement Errors

A theoretical analysis framework has been developed, which is used to evaluate the influence of inaccurate sensor positions, spatial aliasing, and sensor noise on the results of the PWD. Rafaely (2005a) has introduced such a framework for rigid-sphere arrays. First, an arbitrary sound field, which is captured using an open-sphere cardioid array, is considered. The output of a single sensor in

the frequency domain, including the three errors, is written as

$$s(r, \varphi'_q, \vartheta'_q, k) + e_q, \quad (3.47)$$

where k and r are wavenumber and array radius, e_q is the sensor noise, and ϑ'_q and φ'_q denote the inaccurate sensor position. Since, the sound field is sampled the spherical harmonic coefficients $A_{nm}(k)$ are approximated by $\hat{A}_{nm}(k)$. According to Eq. 3.47:

$$\begin{aligned} \hat{A}_{nm}(k) = \frac{1}{b_n(kr)} & \left[\sum_{q=1}^Q w_q \cdot s(r, \varphi'_q, \vartheta'_q, k) \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)} \right. \\ & \left. + \sum_{q=1}^Q w_q \cdot e_q \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)} \right], \end{aligned} \quad (3.48)$$

where Q is the number of sensors, w_q are the quadrature weights, and $b_n(kr)$ is the mode strength, which depends on the array design. The correct sensor positions are denoted by (φ_q, ϑ_q) . The sound field $s(r, \varphi'_q, \vartheta'_q, k)$ can be expressed in terms of the exact coefficients $A_{n'm'}(k)$ using Eq. 3.13. Substituting Eq. 3.13 into Eq. 3.48 yields

$$\begin{aligned} \hat{A}_{nm}(k) = \frac{1}{b_n(kr)} & \left\{ \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} A_{n'm'}(k) \cdot b_{n'}(kr) \right. \\ & \cdot \left[\sum_{q=1}^Q w_q \cdot Y_{n'}^{m'}(\varphi'_q, \vartheta'_q) \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)} \right] + \sum_{q=1}^Q w_q \cdot e_q \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)} \left. \right\}. \end{aligned} \quad (3.49)$$

The term in the square brackets is equivalent to the orthonormality relation of the spherical harmonics $Y_n^m(\phi_q, \theta_q)$ in Eq. (A.19). It can be extended to account for aliasing errors ϵ_a and for positioning errors $\epsilon_{\varphi\vartheta}$ in the following way (Rafaely, 2005a):

$$\begin{aligned} \sum_{q=1}^Q w_q \cdot Y_{n'}^{m'}(\varphi'_q, \vartheta'_q) \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)} = \\ \left\{ \begin{array}{ll} \delta_{n',n} \cdot \delta_{m',m} + \epsilon_{\varphi\vartheta}(n, m, n', m'), & n, n' \leq N_{\max} \\ \epsilon_a(n, m, n', m') + \epsilon_{\varphi\vartheta}(n, m, n', m'), & n \leq N_{\max} < n' \end{array} \right\}, \end{aligned} \quad (3.50)$$

where $\delta_{n',n}$ and $\delta_{m',m}$ are Kronecker delta functions. The maximum level N_{\max} is the highest level of the coefficients $A_{n'm'}(k)$ which can be sampled without spatial aliasing using Q sensor positions. Notice that aliasing errors ϵ_a do not occur in the upper case of Eq. (3.50), since $n' < N_{\max}$. From the Kronecker delta functions it follows that if $\epsilon_{\varphi,\vartheta} = 0$, then $(\varphi'_q, \vartheta'_q)$ and (φ_q, ϑ_q) must be equal, so that $\epsilon_{\varphi,\vartheta}$ represents the positioning error (Rafaely, 2005a). The lower solution of Eq. (3.50) considers a situation where $n' > N_{\max}$, i. e., where spatial aliasing occurs. Since n and n' are different, the term $\delta_{n'n} \cdot \delta_{m'm}$ does not appear. The aliasing error ϵ_a is found from Eq. (3.34), given by

$$\epsilon_a(n, m, n', m') = \sum_{q=1}^Q w_q \cdot Y_{n'}^{m'}(\varphi_q, \vartheta_q) \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)}, \quad n \leq N_{\max} < n'. \quad (3.51)$$

The positioning error $\epsilon_{\varphi,\vartheta}$ is determined by subtracting Eq. (3.50) from Eq. (3.51) (Rafaely, 2005a):

$$\epsilon_{\varphi,\vartheta}(n, m, n', m') = \sum_{q=1}^Q w_q \left[Y_{n'}^{m'}(\varphi'_q, \vartheta'_q) - Y_{n'}^{m'}(\varphi_q, \vartheta_q) \right] \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)}, \quad (3.52)$$

where $n \leq N$, $n' \geq 0$. Inserting (3.50) into (3.49) and separating the summations over n' yields the complete expression to calculate the erroneous spherical

harmonic coefficients.

$$\begin{aligned}
 \hat{A}_{nm}(k) = & \underbrace{\frac{1}{b_n(kr)} \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} A_{n'm'}(k) \cdot b_{n'}(kr) \cdot \delta_{n'n} \cdot \delta_{m'm}}_{A_{nm}^{(s)}(k) - \text{signal contribution}} \\
 & + \underbrace{\frac{1}{b_n(kr)} \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} A_{n'm'}(k) \cdot b_{n'}(kr) \cdot \epsilon_{\varphi\vartheta}(n, m, n', m')}_{A_{nm}^{(\varphi, \vartheta)}(kr) - \text{positioning error}} \\
 & + \underbrace{\frac{1}{b_n(kr)} \sum_{n'=N+1}^{\infty} \sum_{m'=-n'}^{n'} A_{n'm'}(k) \cdot b_{n'}(kr) \cdot \epsilon_a(n, m, n', m')}_{A_{nm}^{(a)}(kr) - \text{aliasing error}} \\
 & + \underbrace{\frac{1}{b_n(kr)} \sum_{q=1}^Q w_q \cdot e_q \cdot \overline{Y_n^m(\varphi_q, \vartheta_q)}}_{A_{nm}^{(e)}(kr) - \text{microphone noise}}.
 \end{aligned} \tag{3.53}$$

The first term in Eq. (3.53) is the error-free contribution to the coefficients $\hat{A}_{nm}(k)$. Due to the product of the Kronecker deltas, the term can be simplified to $A_{nm}(k)$. The last three terms in Eq. (3.53) are the various error signals. They depend on the level n and mode m of the coefficients $\hat{A}_{nm}(k)$, on the product kr , and on the chosen quadrature. Due to the mode strength $b_n(kr)$, the errors also depend on the array design used. The plane wave decomposition is computed by substituting Eq. (3.53) into Eq. (3.40). Each term $A_{nm}^{(\cdot)}(k)$ in Eq. (3.53) yields a separate contribution to the direction weights w_s , which can be written as

$$w_s^{(\cdot)}(\varphi_s, \vartheta_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left[\frac{1}{4\pi i^n} \right] \cdot A_{nm}^{(\cdot)}(k) \cdot Y_n^m(\Omega_s), \tag{3.54}$$

where (φ_s, ϑ_s) is the steering direction of the array and $A_{nm}^{(\cdot)}(kr)$ is $A_{nm}^{(\Omega)}(kr)$, $A_{nm}^{(a)}(kr)$, $A_{nm}^{(e)}(kr)$ or $A_{nm}^{(s)}(kr)$. The effective influence of the various measurement errors on the results of the PWD can be determined by relating the error contributions in Eq. (3.54) to the corresponding signal contribution. The

relative error contributions is obtained as (Rafaely, 2005a):

$$E_{\varphi\vartheta}(kr) = \frac{|w_s^{(\varphi,\vartheta)}(\varphi_s, \vartheta_s, k)|^2}{|w_s^{(s)}(\varphi_s, \vartheta_s, k)|^2}, \quad (3.55a)$$

$$E_a(kr) = \frac{|w_s^{(a)}(\varphi_s, \vartheta_s, k)|^2}{|w_s^{(s)}(\varphi_s, \vartheta_s, k)|^2}, \quad (3.55b)$$

$$E_e(kr) = \frac{|w_s^{(e)}(\varphi_s, \vartheta_s, k)|^2}{|w_s^{(s)}(\varphi_s, \vartheta_s, k)|^2}. \quad (3.55c)$$

In Eq. (3.55), the noise-to-signal ratio (NSR) is used instead of the signal-to-noise ratio (SNR) for comparability to the results of Rafaely (2005a). In the following section the relations presented above are used to explore the various errors. For this purpose, it is common to assume a sound field consisting of a single plane wave. The directional weights $w_s^{(\cdot)}$ and the error contributions $E_{(\cdot)}(kr)$ in Eq. (3.55), respectively, are computed for the direction of arrival (DOA) (φ_0, ϑ_0) of the sound (Rafaely, 2005a). The steering direction (φ_s, ϑ_s) is equivalent to (φ_0, ϑ_s) . Cardioid arrays are studied since a corresponding array is used in the measurements. The rigid sphere design is exhaustively examined by Rafaely (2005a). Six microphone cardioid open sphere arrays are considered and summarized in Table 3.2. To obtain comparable results for the various quadratures, the arrays 1 to 3 are chosen such that they achieve the same N_{\max} , whereas the arrays 4 to 6 provide a similar Q . For all six arrays, microphone noise which is normally distributed with a standard deviation $\sigma_e = 1$ is assumed. This equals a SNR of 0 dB. This value of σ_e is chosen to easily determine the robustness of the PWD against measurement noise. A single ideal plane wave is considered to study the array performance. The DOA $\Omega_0 = (25.7^\circ, 60^\circ)$ agrees with the steering direction Ω_s of the microphone array. The direction Ω_0 is arbitrarily chosen. However, it is shown in the following sections that it has no effect on the presented results.

3.5.2 Spatial Aliasing

As discussed in Section 3.3, a certain amount of spatial aliasing is always present, which is produced by spherical harmonics of higher levels. Nevertheless, the magnitude of a spherical harmonic decays rapidly if the level is higher than kr . Hence, spatial aliasing should be negligible if the sound field satisfies

Array No.	Quadrature	Q	N	N_{\max}	Ω_0 ($^\circ$)
1	Lebedev	14	2	2	(25.7, 60)
2	Chebyshev	20	2	2	(25.7, 60)
3	Gauss-Legendre	12	2	2	(25.7, 60)
4	Lebedev	110	2	8	(25.7, 60)
5	Chebyshev	110	2	5	(25.7, 60)
6	Gauss-Legendre	112	2	7	(25.7, 60)

Table 3.2: Array configurations to evaluate the influences of various measurement errors. Q is the number of microphone positions, N_{\max} is the maximum level of the spherical harmonics which can be sampled without spatial aliasing, and N is the corresponding maximum level used in the calculations.

$kr < N_{\max}$. To verify this behavior, the relative error contribution $E_a(kr)$ of spatial aliasing is computed. The considered array configurations 1 to 6 in Table 3.2 are used to provide comparability. The relative error contribution $E_a(kr)$ is computed using Eq. (3.53), (3.55) and (3.51). The signal contribution $A_{nm}^{(s)}(k)$ and the mode strength $b_n(kr)$ are obtained from Eq. (3.12) and from Eq. (3.19), respectively. The steering directions Ω_s of the microphone arrays are equal to the DOA Ω_0 . Figure 3.12 shows the results for the cardioid design.

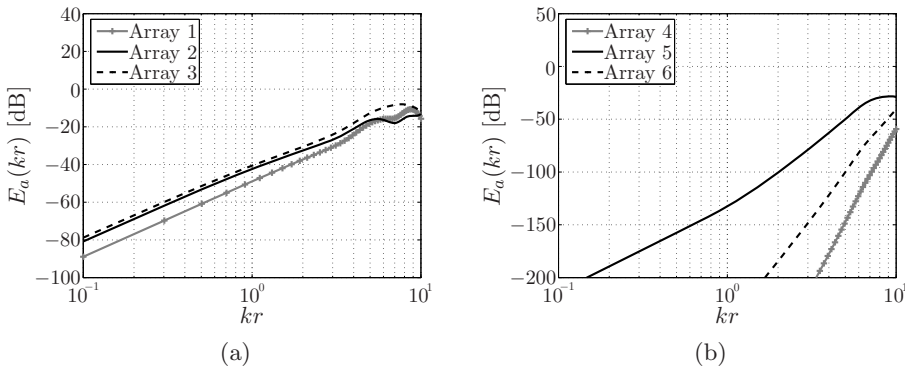


Figure 3.12: Relative error contribution $10 \log_{10}(E_a(kr))$ of spatial aliasing for the open sphere array configurations in Table 3.2 (Thiergart, 2007).

It can be noticed that spatial aliasing increases for increasing kr . This means that it limits the PWD performance when measuring at high frequencies or when using large array radii. If the arrays provide the same N_{\max} , as in (a), then the aliasing errors are similar for all quadratures. This is in contrast to

(b), where the arrays have the same number of microphone positions. In this case, the Lebedev quadrature offers an improved robustness against spatial aliasing. Therefore, this quadrature is recommended when operating in a frequency range, where kr is high. The results in Figure 3.12 show that spatial aliasing should be negligible if $kr < N_{\max}$. In this case, the Lebedev quadrature, for example, is characterized by an error contribution which is below -35 dB in (a) and below -80 dB in (b). The corresponding error contributions of the Chebyshev quadrature are below -30 dB in (a) and below -50 dB in (b). The influence of spatial aliasing depends less on the DOA of the sound. As in the previous section, this assumption was verified by repeating the entire calculations for the DOAs $\Omega_0 = (0^\circ, 90^\circ)$ and $\Omega_0 = (-165^\circ, 22^\circ)$. The results can be found in Appendix B.1.

3.5.3 Microphone Noise

Recent microphone technology can provide a high signal-to-noise ratio (SNR) of typically 90 dB or above (Sennheiser, 2007a), (Sennheiser, 2007b). Nevertheless, microphone noise cannot be disregarded for the plane wave decomposition. In Section 3.4.2 it was demonstrated that uncorrelated noise is significantly amplified if spherical harmonics of high levels n are computed at low values of kr . To determine the influence of microphone noise on the PWD, the relative error contribution $E_e(kr)$ is computed using Eq. (3.53) and Eq. (3.55). The error-free signal contribution $A_{nm}^{(s)}(k)$ in Eq. (3.53) is obtained from Eq. (3.12). The mode strength $b_n(kr)$ follows from Eq. (3.19). The results for $E_e(kr)$ are depicted in Figure 3.13 for $0.1 \leq kr \leq 10$. The presented plots are an average over 50 realizations³ of the microphone noise. The illustrated relative error contribution $E_e(kr)$ is the reciprocal of the so-called white-noise-gain (WNG), which is a measure for the SNR improvement between the array input $s(r, \Omega'_q, k) + e_q$ and the array output $w_s(\Omega_s, k)$ (Bitzer and Simmer, 2001). Therefore, the PWD provides an increased SNR, for regions in Figure 3.13 in which the error $E_e(kr)$ is below 0 dB. In each situation, the highest improvement in SNR is found around $kr = 2$. For both higher and lower kr , it decays significantly. This reflects the reciprocal behavior of the mode strength $b_n(kr)$, which is depicted in Figure 3.4. The omnidirectional open sphere design, which is plotted for comparison in (c) and (d), offers a higher SNR improvement around $kr = N$ compared to the cardioid design in (a) and (b). However, the robust-

³The results are stable already after 30 averages.

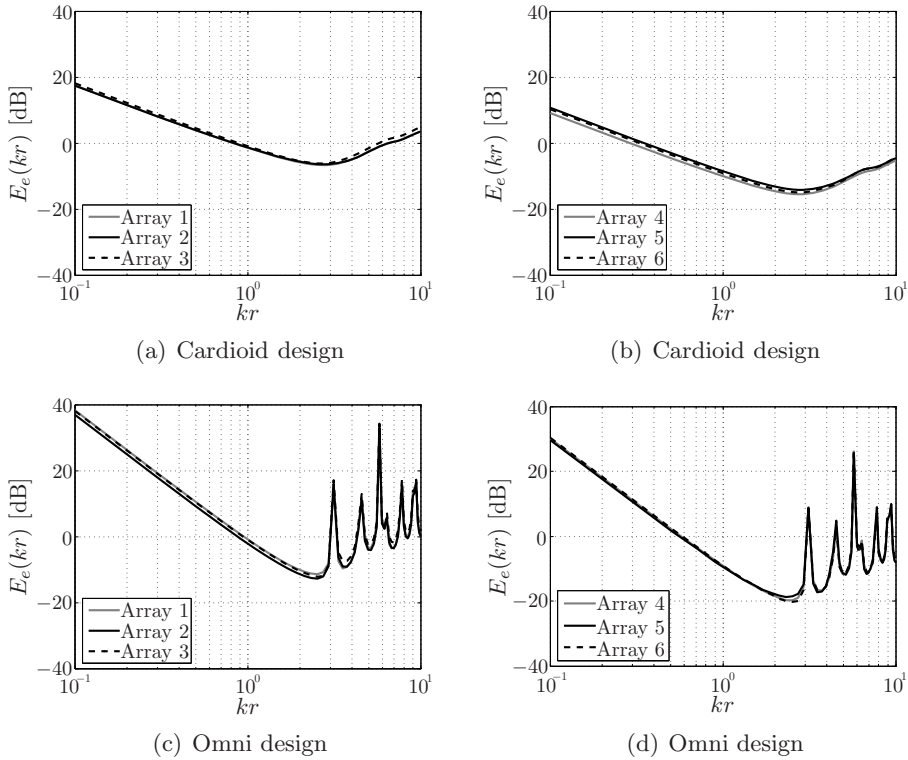


Figure 3.13: Relative error contribution $10 \log_{10}(E_e(kr))$ of microphone noise with a standard deviation $\sigma_e = 1$ for the open sphere array configurations in Table 3.2 (Thiergart, 2007).

ness against noise deteriorates more significantly below and above the error minimum. Moreover, the error contributions in (c) and (d) show several distinct peaks for $kr > N$. Around these kr , microphone noise is enormously amplified. The reason for this is the spherical Bessel function $j_n(kr)$ in Eq. (3.21), which becomes nearly or even exactly zero. Comparing the different array configurations in Figure 3.13, it can be noted that the influence of microphone noise depends only slightly on the chosen quadrature. The higher number of microphone positions in (b) and (d) improves the array robustness significantly against noise compared to (a) and (c). The results in Figure 3.13 are assumed not to depend on the DOA Ω_0 of the sound. To verify this assumption, the entire calculation was repeated for two additional sound fields with very different DOA, namely $\Omega_0 = (0^\circ, 90^\circ)$ and $\Omega_0 = (-165^\circ, 22^\circ)$. The corresponding plots

can be found in the Appendix B.2. It was found that the error contribution is (nearly) independent of Ω_0 .

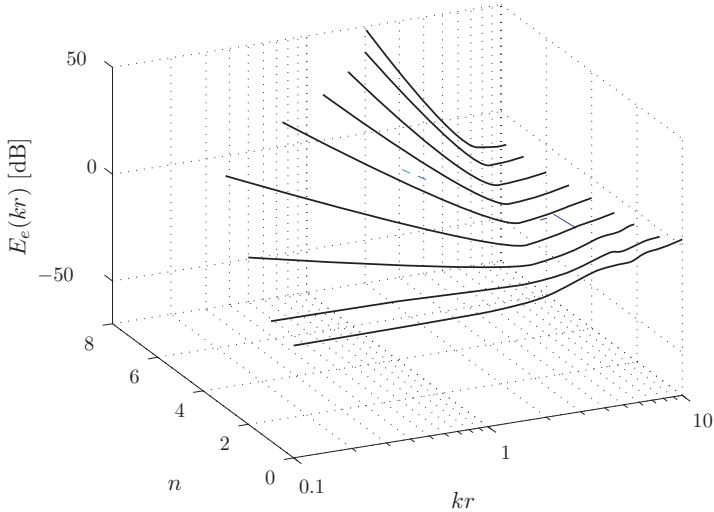


Figure 3.14: Relative error contribution $10 \log_{10}(E_e(kr))$ of microphone noise for the array configuration 4 in Table 3.2 and cardioid sensor elements.

The Lebedev quadrature will often be the sampling scheme of choice, as discussed in Section 3.3. For this reason the Lebedev quadrature of array 4 in Table 3.2 is studied in more detail. Figure 3.14 illustrates the relative error contribution $E_e(kr)$ as a function of kr for the maximum levels $N \in [0, 8]$. The results verify that the error minimum is always located around $kr = N$. Under this constraint, a SNR which is constant over kr is achieved. In general, the error increases less significantly for higher kr than for lower kr . As a result for the practical application, the maximum level can be limited to $N < kr$, for example to reduce the computational effort. It can be concluded from this section that the influence of microphone noise is the lowest if the maximum level is set to $N \approx kr$. In this case, a high (and constant) SNR over a wide frequency range can be achieved. This constraint reduces the resolution of the PWD. As a result, the PWD performance is limited at low frequencies f , or for small array radii r due to a lack of robustness of the array against microphone noise. In other words, a high PWD resolution at low kr involves a high noise amplification. The array robustness against microphone noise does not depend on the chosen quadrature. However, the SNR can be optimized within

the PWD by increasing the number of microphone positions Q .

3.5.4 Positioning Errors of the Sensor Array

In case of virtual sensor arrays as described in Section 3.1.2, positioning errors caused by the mechanical construction can arise. Two possible positioning inaccuracies can be identified:

1. Random positioning errors
2. Constant offset errors.

Both problems are examined in the same way as microphone noise in the previous section. The following results depend on the direction of arrival (DOA) Ω_0 of the sound and on the quadrature considered. Therefore, the error robustness between the various sampling schemes cannot be compared directly. Solving this problem in a generalized way is difficult. A possible way is to repeat the calculations for a specific number of random DOAs and to average the outputs. It is clear that the final result is affected by the distribution of the DOAs. Herein only a single⁴ DOA, as in the previous section, is considered. Nevertheless, the general conclusions from the following results, which are important for the plane wave decomposition, are not affected by Ω_0 or the chosen sampling scheme.

Random Positioning Errors

The random positioning errors are evaluated by computing the relative error contributions $E_{\Omega}(kr)$. The sound field of a single ideal plane wave, with Ω_0 denoting the DOA, is simulated using the six array configurations listed in Table 3.2 on page 62. The steering direction Ω_s of the microphone arrays agrees with the DOA of the sound. Random positioning errors which are normally distributed with a standard deviation $\sigma_{\Omega} = 1^{\circ}$ are assumed. The value of σ_{Ω} is a realistic value for the positioning accuracy of the microphone array used (Schlesinger et al., 2007). The positioning error is added to both components of the direction Ω'_q . While the relative error contribution $E_{\Omega}(kr)$ is determined using Eq. (3.53) and Eq. (3.55), the error-free signal contribution $A_{nm}^{(s)}(k)$ is

⁴This is also conform to the work of (Rafaely, 2005a), in which rigid sphere arrays are examined.

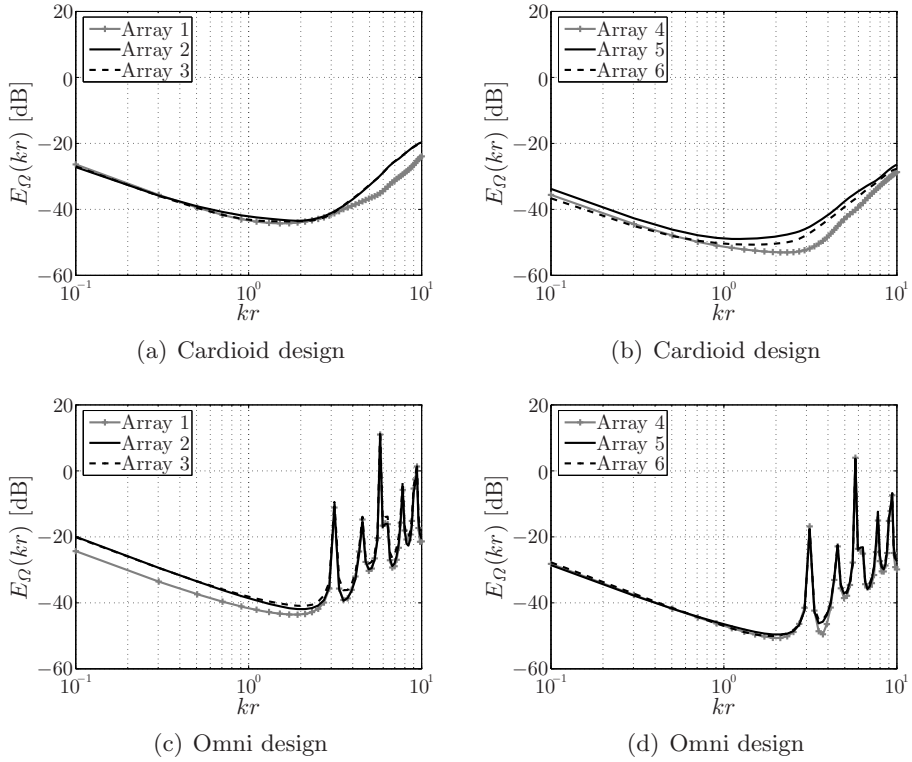


Figure 3.15: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of random microphone positioning errors with a standard deviation $\sigma_{\Omega} = 1^{\circ}$ for the array configurations in Tab. 3.2 using open sphere designs (Thiergart, 2007).

obtained from Eq. (3.12). The mode strength $b_n(kr)$ follows from Eq. (3.21). Figure 3.15(a) and (b) show the results of $E_{\Omega}(kr)$ when using cardioid microphones. These plots present an average over 50 realizations of incorrect microphone positions⁵. It can be noticed that the highest robustness against random positioning errors (i.e., the smallest error contribution) is achieved around $kr = N = 2$. Both a higher and a lower kr leads to increasing errors. This behavior is produced by the $1/b_n(kr)$ term in Eq. (3.53). It can be observed that the characteristic of the error performance is similar for all quadratures. The results in (b) show that the array robustness against positioning errors is significantly higher if a larger number of microphone positions is used. The results for $E_{\Omega}(kr)$ when using open omni sphere arrays are illustrated in Fig-

⁵The results are already stable after 30 realizations.

ure 3.15(c) and (d). The error contributions show a similar behavior as for the cardioid designs in (a) and (b) for small values of kr . This means, the highest array robustness is found around $kr = N = 2$. In contrast to cardioid arrays, several distinct peaks within the error contributions for $kr > N$ can be noticed. The peaks arise exactly for the kr for which the spherical Bessel functions $j_n(kr)$ become zero. Furthermore, the cardioid design gives a slightly better error characteristic than the omni sphere design. A reason could be the lower number of microphones involved in the processing due to the directivity of the single sensors. Because of its benefits for practical applications, the

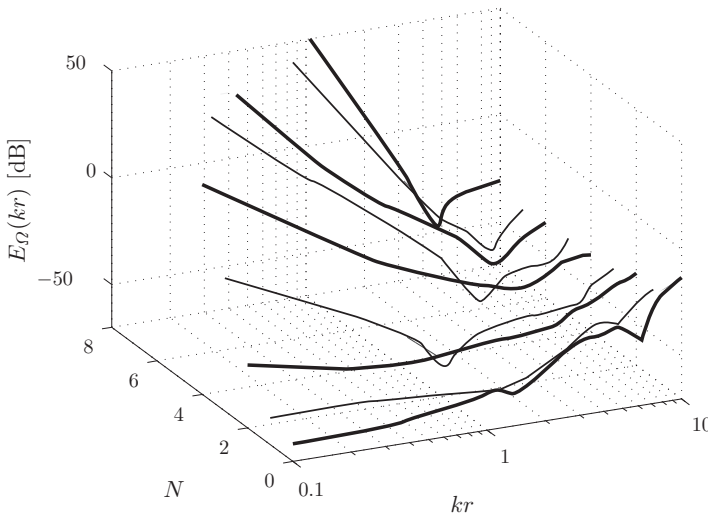


Figure 3.16: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of random microphone positioning errors for the array configuration 4 in Table 3.2, considering a cardioid design.

Lebedev configuration (array 4 in Table 3.2) is studied in more detail. The relative error contributions for the maximum levels $N \in [0, 8]$ are visualized in Figure 3.16. The results for the cardioid design in (a) verify the assumption that the influence of random positioning errors is lowest in case the maximum level is set to $N = \lceil kr \rceil$. Nevertheless, it can be observed that the error contribution increases under this constraint. In other words, the array robustness decreases for both larger array radii or higher frequencies. The decrease in robustness at higher frequencies can be explained by a shorter wavelength of sound, which yields an increased risk of sampling an incorrect phase. The same

problem can arise when using a larger array radius, which leads to a larger offset of the individual microphones. The results of this section show that the influence of random positioning errors on the results of the plane wave decomposition is similar to the behavior of microphone noise. A minimum error can be achieved if a maximum level $N \approx kr$ is used within the calculations of the PWD. However, this lowers the resolution of the PWD at low values of kr and increases the computational effort at high values of kr . The general conclusions of this section can be applied to all quadratures considered. To prove them, the calculations in Figure 3.15 were repeated for the sound arrival directions $\Omega_0 = (0^\circ, 90^\circ)$ and $\Omega_0 = (-165^\circ, 22^\circ)$ (see Appendix B.3), which verify the outcomes of this section.

Offset Positioning Errors

The influence of constant offset positioning errors is analyzed in the following. These errors can be divided into horizontal offset errors in azimuth direction φ and vertical offset errors in co-elevation direction ϑ . The former error occurs, for example, if the horizontal alignment of the microphone array is incorrect. The latter errors can arise if the vertical step motor is over-strained due to the weight of the robot arm. Horizontal offset errors are less critical since they only rotate the coordinate system in the azimuthal direction. However, vertical offset errors are more critical. The microphone angles in azimuth direction vary in the range $-\pi \leq \varphi_q \leq \pi$, whereas the range for the angles in co-elevation direction is $0 \leq \vartheta_q \leq \pi$. A constant offset error added to ϑ_q does not simply rotate the coordinate system, but it leads to denser microphone positions near one pole of the measurement sphere (see Figure 3.17). The vertical offset positioning errors are evaluated in the same way as the random positioning errors (i. e., by computing the contribution of the offset errors to the direction weights w_s). To provide comparable results, the same array configurations are considered as in the previous sections. A plane wave arriving from Ω_0 is assumed, which corresponds to the steering direction Ω_s of the microphone arrays. A constant angle $\gamma_\vartheta = 1^\circ$ is added to the co-elevation ϑ_q of each microphone position to simulate the offset errors. This value is chosen to provide results comparable to the previous section. To examine the influence of the offset positioning error, the relative error contribution $E_\Omega(kr)$ is computed. However, a constant angle γ_ϑ is added only to the co-elevation components of the microphone positions Ω'_q . Figure 3.18(a) and (b) show the results for the cardioid microphones. Similarly to the random positioning

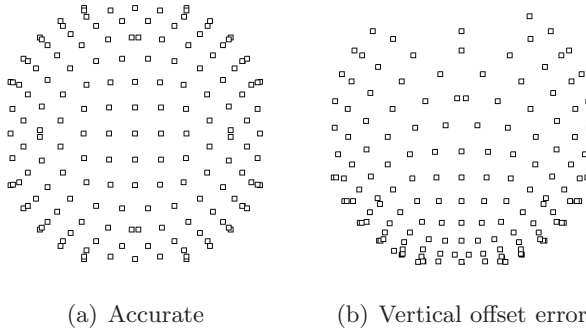


Figure 3.17: Plot (a) shows an example for accurate microphone positions. Plot (b) shows inaccurate positions as a result of adding a (unrealistically) large constant vertical offset error, which yields a denser sampling near the south pole of the sphere.

errors, the minimum of the error contributions is found around $kr = N = 2$. Comparing Figure 3.18(a) with (b), it can be observed that the influence of offset errors is not reduced if the number of microphone positions is increased. As in the previous sections, the Lebedev quadrature with array 4 in Table 3.2 is studied in detail. The relative error contribution $E_{\Omega}(kr)$ for both array designs are depicted in Figure 3.19 for the maximum levels $N \in [0, 8]$. The error robustness against offset errors is generally improved if N is low. This is in contrast to the positioning errors in Figure 3.16 and the microphone noise in Figure 3.14, in which the error minimum is always found around $kr = N$. The array robustness deteriorates significantly if $N > kr$, which agrees with the results for the random positioning errors and the microphone noise. Therefore, a reduction of N might be necessary in practice. Referring to Section 3.4.1, this limits the resolution of the PWD, particularly at low frequencies or for small array radii, where kr is low.

The results of this section show that the influences of random positioning errors depend only slightly on the maximum level N , as long as $N < kr$. However, N must be reduced at low kr to avoid a poor error robustness. Furthermore, the robustness of the array can be improved by increasing the number of microphone positions. The results of this section have been verified by repeating the calculations in Figure 3.18 for the DOAs $\Omega_0 = (0^\circ, 90^\circ)$ and $\Omega_0 = (-165^\circ, 22^\circ)$ (see Appendix B.3).

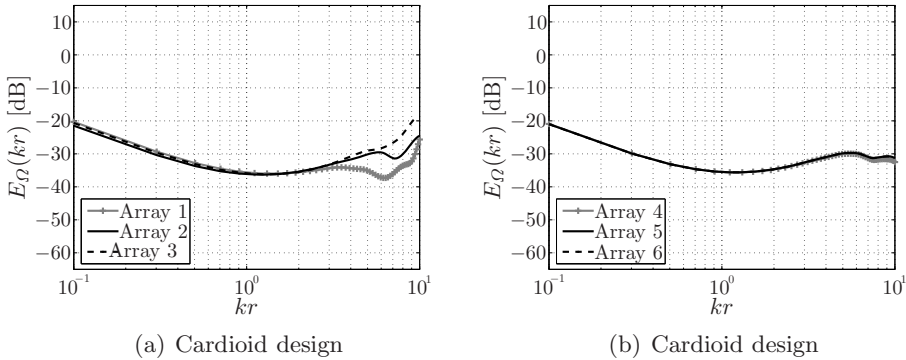


Figure 3.18: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of vertical offset positioning errors for the array configurations in Table 3.2 (Thiergart, 2007).

3.5.5 Non-ideal Microphone Characteristics

Cardioid microphones provide a solution to the numerical problems of the open sphere design because no singularities arise in the mode strength. However, ideal (frequency-independent) microphone characteristics cannot be assumed in practice. To demonstrate a non-ideal cardioid microphone characteristic, Figure 3.20 shows the polar diagrams of a Sennheiser⁶ microphone (Sennheiser, 2007c) in contrast to an ideal cardioid characteristic. Especially at higher frequencies the non-ideal⁷ characteristic differs significantly. In the following section, the influences of a non-ideal first-order characteristic on the PWD are analyzed in a similar manner as for the positioning errors and the microphone noise. To evaluate these influences, a mathematical description is derived and the results are presented.

Error Functions for Non-ideal Microphone Characteristics

A pragmatic approach to compute the error contribution of a non-ideal microphone characteristic to the PWD is proposed in this section. A sound field consisting of a single plane wave with a DOA Ω_0 is assumed. To simulate realistic cardioid microphones, the MKH 8040 characteristic in Figure 3.20 is

⁶Sennheiser characteristics are considered, since the polar diagrams are available on a high resolution on <http://www.sennheiser.de>

⁷Non-ideal named the realistic case of a practical realization.

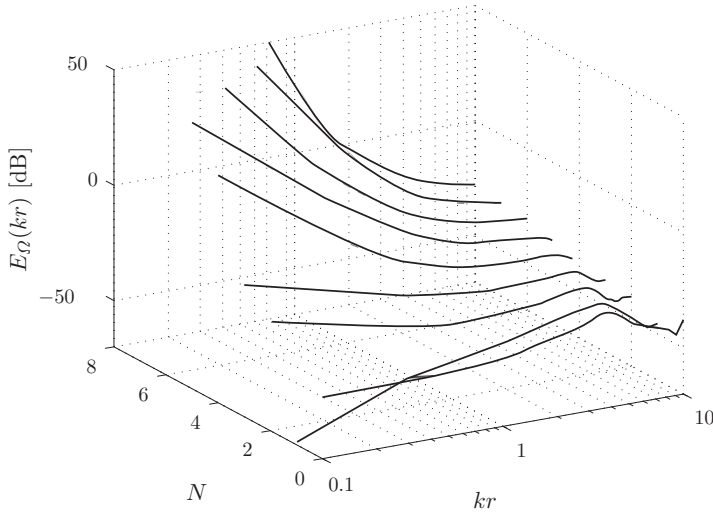


Figure 3.19: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of vertical offset positioning errors for the array configuration 4 in Table 3.2, considering a cardioid design (Thiergart, 2007).

used, which is available for $125 \text{ Hz} \leq f \leq 16 \text{ kHz}$. To avoid the occurrence of additional errors when studying such a wide frequency range, it is crucial to ensure that no spatial aliasing occurs in the simulation. To compute the error contribution of a non-ideal microphone characteristic, the following steps are performed:

1. Determine the ideal coefficients $A_{nm}(k)$ of the considered sound field using Eq. 3.12. The coefficients are computed for a maximum level N .
2. Use the results to compute the sound pressure $p(r, \Omega_q, k)$ for each microphone position Ω_q . As a result, the pressure field is always level-limited to N . Thus, a frequency-independent constant number of microphone positions can be used for all kr without producing spatial aliasing.
3. Weight⁸ the pressures with the ideal and non-ideal microphone characteristic in Figure 3.20 to obtain the first-order microphone responses. Depending on the frequency, the non-ideal characteristic is linearly interpolated.

⁸The weighting factor depends on the angle between Ω_q and Ω_0 .

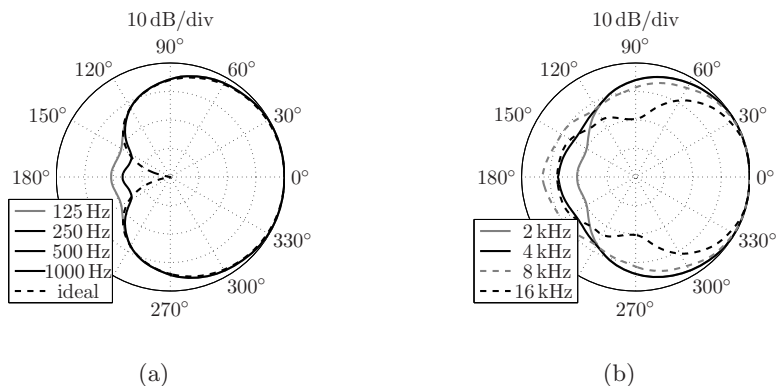


Figure 3.20: Polar diagrams of Sennheiser MKH 8040 at different frequencies (Sennheiser, 2007c) and ideal characteristic.

4. Determine the corresponding ideal and non-ideal coefficients $A_{nm}(k)$ and $\hat{A}_{nm}(k)$, respectively, using Eq. (3.53).
5. Compute the corresponding directional weights $w_s(\Omega_s, k)$ and $\hat{w}_s(\Omega_s, k)$ from the SFT approach.

Subsequently, the root mean squared error (RMSE) is introduced to compute the absolute error contribution of a non-ideal cardioid microphone characteristic as

$$w_s^{(\Theta)}(\Omega_s, k) = \sqrt{\text{E}[|w_s(\Omega_s, k) - \hat{w}_s(\Omega_s, k)|^2]}, \quad (3.56)$$

where $\text{E}[\cdot]$ denotes the average value. Similarly to the approach in Eq. (3.55), the relative error contribution can be obtained by relating the absolute error contribution $w_s^{(\Theta)}$ to the error-free signal contribution \hat{w}_s :

$$E_{\Theta}(k) = \frac{|w_s^{(\Theta)}(\Omega_s, k)|^2}{|w_s(\Omega_s, k)|^2}. \quad (3.57)$$

Eq. (3.57) is used in the following section to evaluate the influence of a non-ideal microphone characteristic for different kr .

Non-ideal Cardioid Microphones

This part deals with the influence of a non-ideal cardioid microphone characteristic on the results of a PWD. Array configurations 7–12 in Table 3.3 are

Array No.	Quadrature	Q	N	N_{\max}	r [m]	f [Hz]
7	Lebedev	110	8	8	1.5	[125, 16 k]
8	Gauss-Legendre	144	8	8	1.5	[125, 16 k]
9	Chebyshev	272	8	8	1.5	[125, 16 k]
10	Lebedev	770	8	23	1.5	[125, 16 k]
11	Gauss-Legendre	760	8	19	1.5	[125, 16 k]
12	Chebyshev	812	8	14	1.5	[125, 16 k]
13	Lebedev	110	8	8	[0.19, 24]	1 k
14	Gauss-Legendre	144	8	8	[0.19, 24]	1 k
15	Chebyshev	272	8	8	[0.19, 24]	1 k
16	Lebedev	770	8	23	[0.19, 24]	1 k
17	Gauss-Legendre	760	8	19	[0.19, 24]	1 k
18	Chebyshev	812	8	14	[0.19, 24]	1 k

Table 3.3: Array configurations to study influences of a non-ideal microphone characteristic, where r is the array radius and f is the operating frequency range. The frequency range corresponds to the frequencies for which the Sennheiser MKH 8040 characteristic is available. The number of microphone positions Q allows the sampling of a sound field up to a maximum level N_{\max} without spatial aliasing. The symbol N denotes the highest level within the calculations.

considered first. The values of Q and N are increased compared to the previous sections since the kr -range is higher. To achieve comparability, the arrays 7–9 are chosen such that they provide the same N_{\max} , whereas the arrays 10–12 have a similar Q . Based on the previous sections, a sound field consisting of a single plane wave with DOA $\Omega_0 = (25.7^\circ, 60^\circ)$ is considered. Furthermore, the sound field is level-limited to N to avoid spatial aliasing. To verify the influence of the DOA Ω_0 and the chosen quadrature, the following calculations are repeated for the DOAs $\Omega_0^{(2)} = (0^\circ, 90^\circ)$ and $\Omega_0^{(3)} = (-165^\circ, 22^\circ)$. The results are depicted in Appendix B.4. The relative error contribution $E_\Theta(kr)$ is calculated, the steering direction Ω_s is equal to the DOA Ω_0 . The results for the array configurations 7–12 are depicted in Figure 3.21. Comparing the plots, it can be noticed that the error contribution $E_\Theta(kr)$ depends neither on the number of microphones Q , nor on the chosen quadrature. Furthermore, the error contribution does not increase for higher values of kr , although the discrepancy between the ideal and non-ideal microphone characteristic is larger at higher frequencies. In this special case the influence of a non-ideal microphone characteristic on the results of the PWD is low. However, this is

only true for sound fields consisting of a single plane wave and if the steering direction Ω_s of the array corresponds to the DOA Ω_0 of the sound. The increasing error contribution $E_\Theta(kr)$ towards smaller kr can be explained with an array robustness, which is generally reduced for $kr < N$, due to a low mode strength $b_n(kr)$. The fluctuations present at higher kr are caused by the level restriction of the sound field. The outputs for the two additional DOAs $\Omega_0^{(2)}$ and $\Omega_0^{(3)}$ verify these conclusions. This shows that the influence of a non-ideal microphone characteristic does not depend on the quadrature or the direction of arrival of the sound. To confirm the results, the error contribution $E_\Theta(kr)$

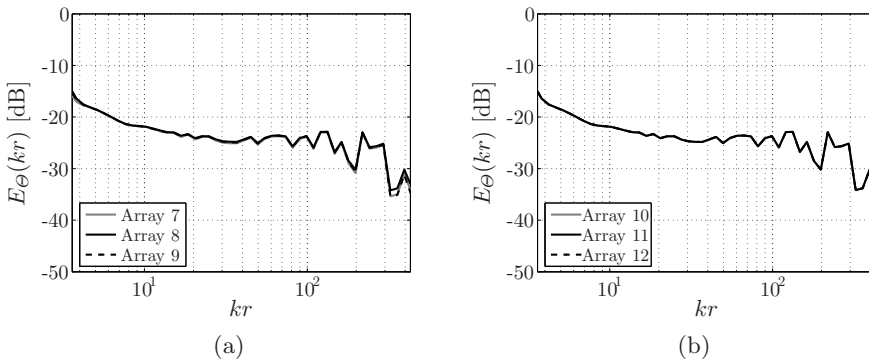


Figure 3.21: Relative error contribution $10 \log_{10} E_\Theta(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a constant radius $r = 1.5$ m (Thiergart, 2007).

is computed for a constant frequency, but for different array radii. In doing so, the influences of a varying realistic microphone characteristic are neglected. The array configurations 13–18 in Table 3.3 are used, which provide the same kr -range as the arrays 7–12. Moreover, the maximum level N and the number of microphones Q remain untouched to achieve comparability. However, notice that a radius of $r = 24$ m is not realizable in practice. The results for $E_\Theta(kr)$ are illustrated in Figure 3.22. It can be observed that the error contribution $E_\Theta(kr)$ is nearly constant for $kr > N$. In contrast, the error increases for $kr < N$, although the discrepancy between the ideal and non-ideal microphone characteristic does not. As a next step, a sound field consisting of a single plane wave arriving from $\Omega_0 = (0^\circ, 90^\circ)$ is considered. The sound field is limited to level $N = 8$ to avoid spatial aliasing. For the simulation Configuration 7 from Table 3.3 is used. Now three different first-order directivities are considered (see Figure 3.23). The erroneous directional weights $\hat{w}_s(kr)$ are computed and

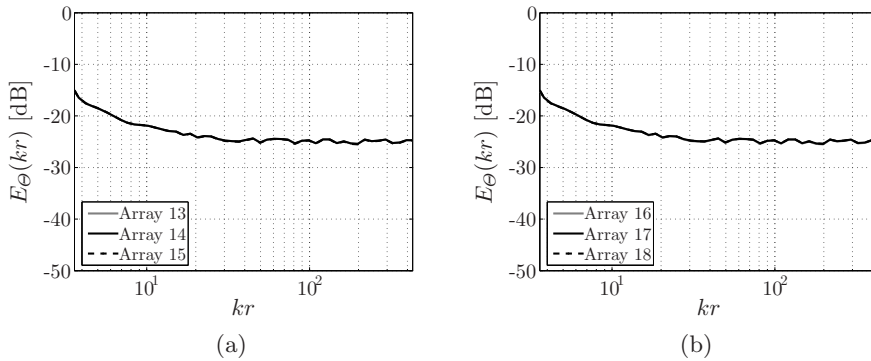


Figure 3.22: Relative error contribution $10 \log_{10} E_{\Theta}(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a constant frequency $f = 1$ kHz (Thiergart, 2007).

the results for a co-elevation of $\vartheta_s = 90^\circ$ are plotted against the azimuth φ_s to obtain the horizontal directional response of the array. The plots in the lower row in Figure 3.23 show the resulting directional responses of the PWD (i. e., the polar diagrams of the directional weights $\hat{w}_s(kr)$). For reference, the plots (a) and (e) illustrate the simulation when using ideal cardioid microphones. In all plots, the sound arrives from the right. Comparing the non-ideal PWD results in (f)–(h) with the reference in (e), it can be noticed that all main lobes (i. e., the lobes which point to the right) are nearly unaffected by the corresponding non-ideal microphone characteristic. This explains the low error contributions $E_{\Theta}(kr)$ in the previous subsection, which are achieved because both the steering direction of the microphone array and the DOA of the sound are equal. However, it is visible in Figure 3.23 that a non-ideal characteristic increases the sensitivity toward the back of the array enormously.

3.6 Bandwidth Extension Using Multi-radii Design

In practice, $A_{nm}(k)$ can only be computed correctly for $n \leq N_{max}$, where N_{max} represents the maximum order of robust calculation. This is a consequence of the fact that the mode strength $b_n(kr)$ becomes almost zero for increasing n at constant values of kr . This highly amplifies measurement errors which can corrupt the results (Rafaely, 2005a). Hence, the first sum in Eq. (3.40) must be limited to N_{max} , whereas a small N leads to a low spatial resolution of the

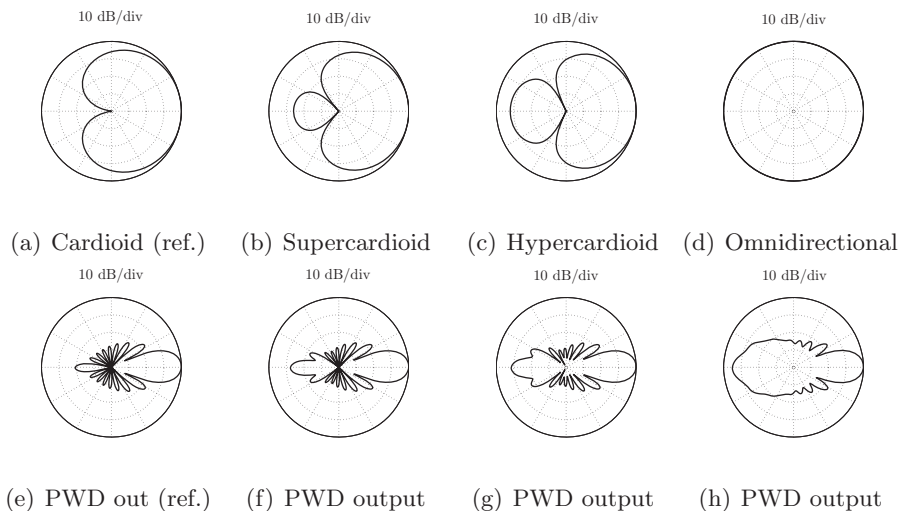


Figure 3.23: The upper plots show the characteristics of the microphones of a first-order array in dB scale. The lower plots show the resulting directional responses of the PWD for $N = 8$. The plots (a) and (e) are the ideal reference (Thiergart, 2007).

PWD (Rafaely, 2004). Since N is lower for smaller kr , a high spatial resolution at low frequencies requires a large array radius, which in turn leads to stronger spatial aliasing at higher frequencies. To overcome this problem, an array featuring two microphones at two different radii r_1 and r_2 can be used. The large radius measurement on r_2 is used to determine the coefficients $A_{nm}(k)$ below the *transition frequency* f_c and the small radius measurement on r_1 is used for frequencies above the transition frequency. The array parameters are a compromise between constraints of a practical realization of a virtual open sphere sensor array and the accuracy of the PWD. Discrete switching between two radii at a single frequency can lead to other undesirable artifacts. At the transition frequency, the spatial resolution changes according to the N_{max} used. This leads to a distribution of the energy over an increased number of direction weights. Figure 3.24(c) shows the directional weights for a sound field consisting of a single plane wave. A single co-elevation angle which corresponds to the direction of arrival of the plane wave is plotted. The transition frequency is set to $f_c = 1900$ Hz. The abrupt change in resolution in direction of 0° can be noticed. If a single direction is extracted for post processing (e. g., in adaptation to binaural reproduction), the discontinuity in the spectrum becomes audible as a sinusoid component, due to the step in the magnitude of the resulting

frequency response. A spectrogram of the DOA of the plane wave is depicted in Figure 3.24(d). The sinusoid at the transition frequency can clearly be identified. To avoid such artifacts, two filters $H_1(k)$ and $H_2(k)$ can be used for r_1 and r_2 , respectively. These filters should be designed such that

$$1 = |H_1(k)| + |H_2(k)| \quad (3.58)$$

can be guaranteed for all k . The filter can be applied to the directional weights obtained from the PWD according to Eq. (3.40). The broadband direction weights $w_s(\varphi_s, \vartheta_s, k)$ are given by

$$w_s(\varphi_s, \vartheta_s, k) = H_1(k)w_{s1}(\varphi_s, \vartheta_s, k) + w_{s2}(\varphi_s, \vartheta_s, k)H_2(k), \quad (3.59)$$

where (φ_s, ϑ_s) denotes the steering direction. The filter can be applied directly to the coefficients $A_{nm}(k)$ when calculating the PWD given by Eq. (3.40). Hence

$$w_{s\langle\cdot\rangle}(\varphi_s, \vartheta_s, k) = H_{\langle\cdot\rangle}(k) \cdot \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{1}{4\pi i^n} A_{nm\langle\cdot\rangle}(k) Y_n^m(\varphi_s, \vartheta_s), \quad (3.60)$$

where $\langle\cdot\rangle$ denotes the radius used. Alternatively, the same result can be achieved by filtering the the coefficients measured at the two different radii:

$$A_{nm}(k) = H_1(k)A_{nm1}(k) + H_2(k)A_{nm2}(k). \quad (3.61)$$

To avoid aliasing and guarantee a stable calculation of the coefficients, the maximum level N_{max} used for calculation is set to $N_{max} = kr$. This can be interpreted as a filtering of the coefficients, which can be included in the filter $H_{\langle\cdot\rangle nm}(k)$:

$$H_{\langle\cdot\rangle nm}(k) = \begin{cases} H_{\langle\cdot\rangle}(k) & \forall n \leq \lceil kr_{\langle\cdot\rangle} \rceil \\ 0 & \forall n > \lceil kr_{\langle\cdot\rangle} \rceil, \end{cases} \quad (3.62)$$

where $r_{\langle\cdot\rangle}$ is the radius for which the filter is calculated. This approach is only one of several possible solutions. Alternatively, a Wiener filter could be used to optimize the signal energy contained in the higher levels. A similar approach was proposed by Kuntz and Rabenstein (2008) and studied by Kuntz (2009) for the use in circular array processing.

3.6.1 Comparison of Multi-radii Designs

Two filters have been used for the application of the multi-radii design discussed in the following sections. The first filter is optimized for the highest possible resolution. This is applied during the extraction of reflections. The filter for the coefficients of the two radii is given by Eq. (3.62) where $\langle \cdot \rangle$ indicates the radius 1 or 2, with

$$H_1(k) = \begin{cases} 1 & \forall k \leq k_c \\ 0 & \forall k > k_c \end{cases} \quad (3.63)$$

$$H_2(k) = \begin{cases} 0 & \forall k \leq k_c \\ 1 & \forall k > k_c. \end{cases} \quad (3.64)$$

The frequency range is given by $k_c = 2\pi f_c/c$, where f_c denotes the transition frequency of the two radii. The second filter minimizes the artifacts in auralization at the cost of the resolution:

$$H_{1nm}(k) = \begin{cases} H_1(k) & \forall n \leq \min([\lceil kr_1 \rceil \lceil k_c r_2 \rceil]) \\ 0 & \forall n > \lceil kr_{\langle \cdot \rangle} \rceil, \end{cases} \quad (3.65)$$

$$H_{2nm}(k) = \begin{cases} H_2(k) & \forall n \leq \lceil kr_2 \rceil \\ 0 & \forall n > \lceil kr_{\langle \cdot \rangle} \rceil. \end{cases} \quad (3.66)$$

The resolution is adapted by limiting the maximum n used for the calculation for the large array to the resolution of the small radius at the transition frequency. To smooth the transition between the coefficients, a cross fade is applied by

$$H_1(k) = \begin{cases} 1 & \forall k \leq k_c - \alpha \\ 0.5 + 0.5 \cos\left(\pi c \frac{k - (k_c - \alpha)}{(k_c + \alpha) - (k_c - \alpha)}\right) & \forall k_c - \alpha < k \leq k_c + \alpha \\ 0 & \forall k > k_c + \alpha \end{cases} \quad (3.67)$$

$$H_2(k) = \begin{cases} 0 & \forall k \leq k_c - \alpha \\ 0.5 - 0.5 \cos\left(\pi c \frac{k - (k_c - \alpha)}{(k_c + \alpha) - (k_c - \alpha)}\right) & \forall k_c - \alpha < k \leq k_c + \alpha \\ 1 & \forall k > k_c + \alpha. \end{cases} \quad (3.68)$$

The two methods are compared to a single radius array design. This is done by simulations of a single plane wave with a DOA of $(0^\circ, 90^\circ)$. The responses

of the array are depicted in Figure 3.24. The simulations are limited to a frequency range of 0 Hz to 3500 Hz. The transition frequency of the two radii was set to 2100 Hz to avoid aliasing in the coefficients of the larger radius. The used radii are 0.28 m and 1 m. The plot of a single plane of directional weights shows the resolution of the different designs. In the upper row the results for a single radius of $r_2 = 0.28$ m are given. The resolution deteriorates in the frequency range below 2000 Hz. The hard switch of the two radii is given in the middle row. The difference in resolution is clearly visible at the transition frequency. The consequence of this hard switch can be seen in the spectrogram on the right hand side, which was calculated for the direction weights of the DOA of the plane wave. At the transition frequency f_c , a strong sinusoidal component is observed. If a resolution adaptation and a fade of the coefficients is used, the resolution is constant below the transition frequency where n is set to a constant value. No sinusoidal component arises at the transition frequency because fading and the resolution adaptation result in a smooth frequency response. It is important to note that the adaptation of the resolution is sufficient to avoid the sinusoid in the ideal case. In the non-ideal case the fade is required. In this case, the theoretical resolution at the transition point is no longer achieved.

3.7 Quantitative Analysis of Measurement Errors

Based on the theoretical findings in the previous section, a prototype array system was used to evaluate the possibilities of a virtual open sphere cardioid microphone array. The first prototype was realized by Schlesinger et al. (2007). In a second iteration, the array was optimized in terms of measurement accuracy. This system is used in order to verify the theoretical analysis. In parallel, simulations were conducted using the same parameters as the prototype array system summarized in Table 3.4. A picture of the dual radius

Q_{Leb}	f_{max}	r_1	f_1	r_2	f_c	mic type
2030	20 kHz	0.28 m	7.4 kHz	1 m	2.1 kHz	cardioid

Table 3.4: Open sphere array configuration used in the measurement and simulation.

sensor array used is given in Figure 3.25(a). The geometrical setup of the simulation and the measurement consisting of a single sound source is depicted in Figure 3.25. The measurement was carried out in the anechoic measurement

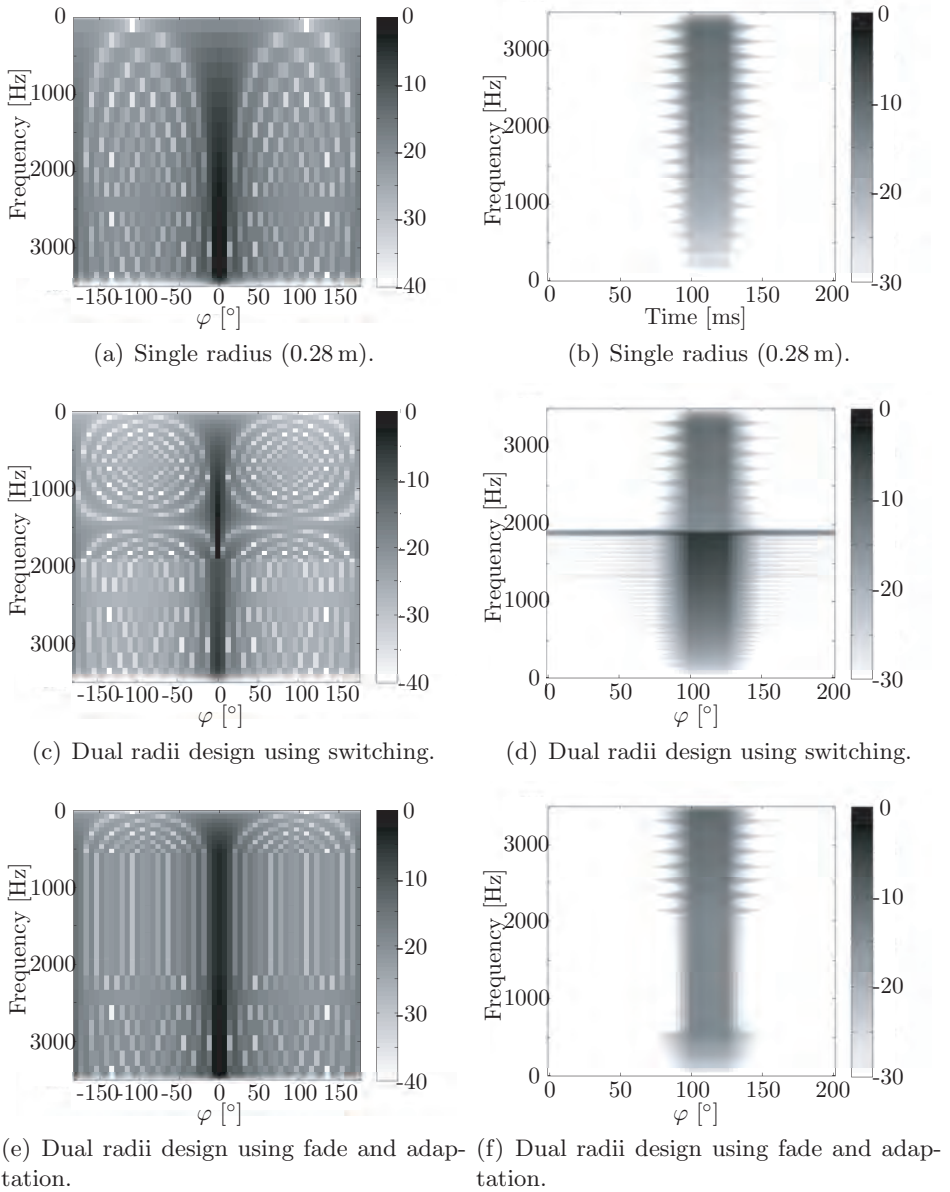


Figure 3.24: Comparison of different array designs using an ideal simulation of a single plane wave. The left column shows the direction weights ($10\log_{10} |w_s(\varphi, 90^\circ, \omega)|$) for the co-elevation corresponding to the direction of arrival of the plane wave. The right column shows spectrograms of the direction weights in the direction of arrival of the plane wave.

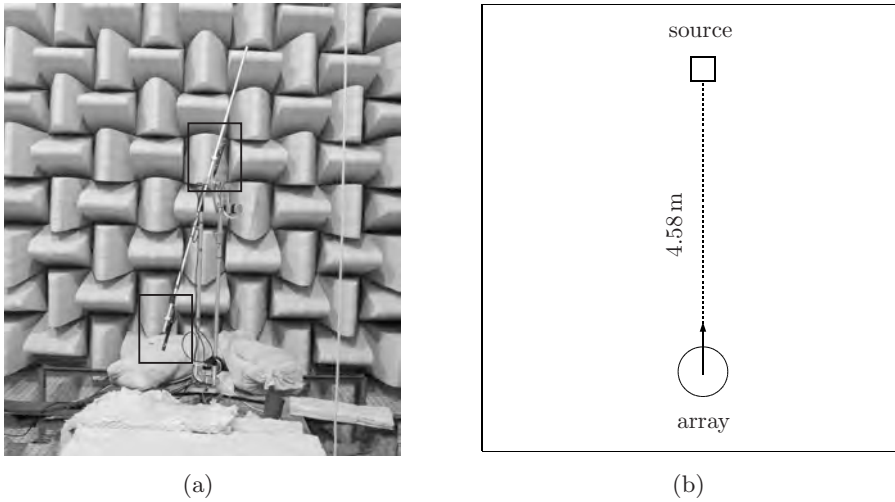


Figure 3.25: (a) The anechoic measurement room of Delft University of Technology and the spherical sensor array used. The microphones, marked by the rectangles, are mounted on both sides of the robot arm for a better balance. (b) The geometry for the simulation and measurement (top view) (Thiergart, 2007).

room at Delft University of Technology. The simulation includes the following measurement errors as well as a realistic cardioid sensor characteristic. The values are based on the analysis of the sensor array used for the measurements given in (Schlesinger et al., 2007):

- Zero mean Gaussian distributed random positioning error in azimuth with standard deviation $\sigma_1 = 1^\circ$ and in co-elevation with $\sigma_2 = 4^\circ$,
- Constant offset positioning error in co-elevation: 3° ,
- Normally distributed white noise signal e_q resulting in a 80 dB SNR.

To compute the impulse responses for each sensor, first a flat frequency spectrum is generated. The phase is adjusted according to the travel time of the sound from the omni-directional source to the sensor position. A $1/d$ decrease of the sound pressure is included, where d is the traveling distance of the sound. The spherical harmonic decomposition and plane wave decomposition are computed for several steering directions Ω_s . The maximum order is set to $N = \lceil kr \rceil$ to ensure the highest error robustness, where r is either r_1 or r_2 depending on the frequency. Furthermore, N is restricted to $N \leq 38$, which

takes effect for $f > f_1$, to limit the computational effort. The normalized directivity function $w_s(\varphi, \vartheta, k)$, determined from the simulations and the measurement, is depicted in Figure 3.26 for the transition frequency $f_c = 2.1$ kHz. The peak in each plot represents the direct sound. For the smaller radius r_1 (Figure 3.26(b), 3.26(d)) nearly no influence of the measurement error is visible at this frequency. The measurement error is clearly visible for the larger radius r_1 (Figure 3.26(a), 3.26(c)). The Figure 3.27 shows the results ob-

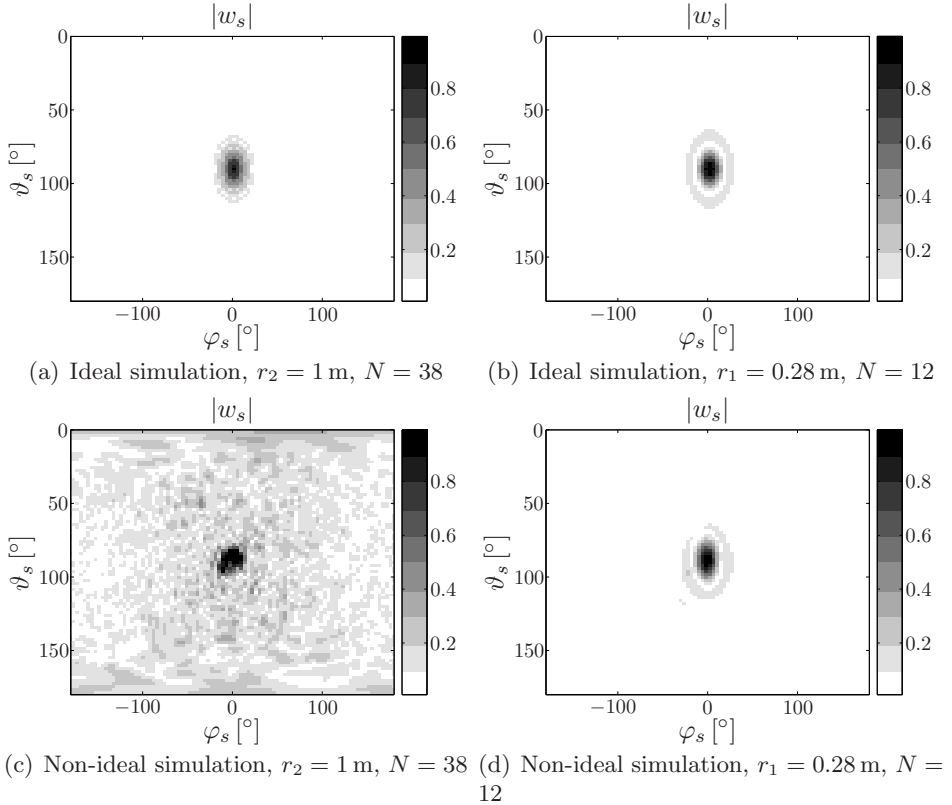


Figure 3.26: Direction weights $|w_s|$ at $f = 2.1$ kHz for the simulation setup in Figure 3.25(b). The results are normalized for a unit maximum magnitude (Thiergart, 2007).

tained by the measurement. It was found that the performance of the real array is better than expected using the assumed measurement errors within the simulations. Figure 3.28(a) shows the directivity function $w_s(\varphi, \vartheta, k)$ determined with the error-free simulation for the horizontal plane $\Omega_s = (\varphi, 90^\circ)$

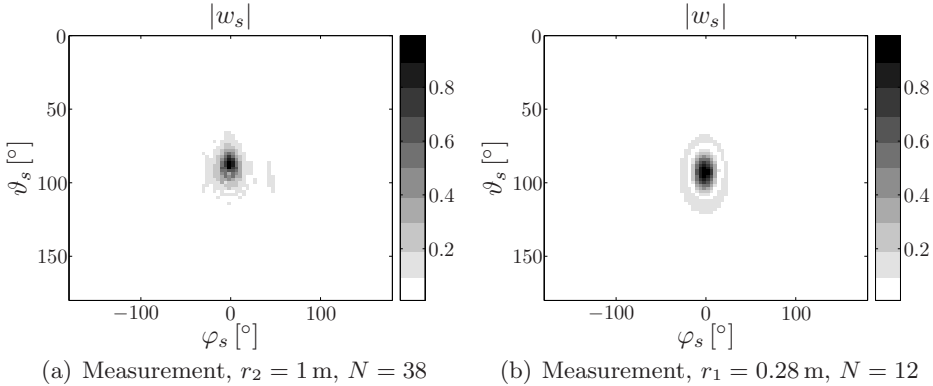


Figure 3.27: Direction weights $|w_s|$ at $f = 2.1$ kHz for the measurement setup in Figure 3.25(b). The results are normalized for a unit maximum magnitude (Thiergart, 2007).

as a function of f . Figure 3.28(b) shows the results of a simulation including simulated measurement errors. Spatial aliasing is present above $f = 12$ kHz. The theoretical PWD resolution (Eq. (3.45)) is marked by the dashed lines in Figure 3.28(a) and specifies the area in which the direct sound energy is mainly concentrated (Rafaely, 2004). Figure 3.29(a) shows the directional weights w_s as a function of f for the *vertical* plane between source and array (i. e., for all steering directions with an azimuth $\varphi_s = -2^\circ$). An undesired microphone offset can clearly be seen. The PWD results below the transition frequency imply that the microphone on r_2 is tilted upwards. This leads to distortions within the directional weights, which is slightly visible in Figure 3.27(a). The equalized directional weights (compensated for the transfer function (TF) of the measurement chain) w_s for the *horizontal* plane between source and array are visualized in Figure 3.29(b). Below the transition frequency, marked by the dotted line, the results of r_2 for a co-elevation $\vartheta_s = 84^\circ$ are plotted. Above this frequency, the results of r_1 for $\vartheta_s = 88^\circ$ are depicted. By this, the vertical offset between both microphones is compensated. As a result, the correct transfer function of the direct sound under an azimuth $\varphi_s = -2^\circ$ is obtained. The results in Figure 3.29(b) show a strong similarity to the ideal simulation results in Figure 3.28(a). The array robustness decreases only slightly at higher frequencies in the measurement case. As expected, spatial aliasing arises above approximately $f = 11$ kHz. In contrast to the non-ideal simulation in Figure 3.28(b), the aliasing energy is more concentrated in specific regions. The transfer function for the DOA of the direct sound in Figure 3.29(b) is de-

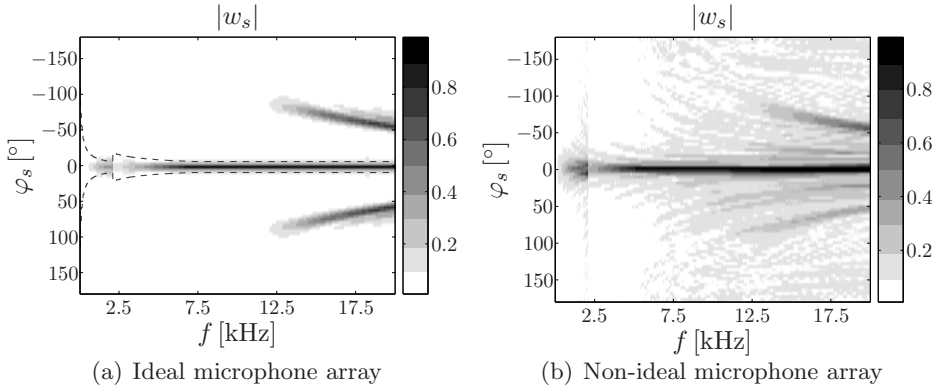


Figure 3.28: Directional weights $|w_s|$ for the simulation as a function of f for a co-elevation $\vartheta_s = 90^\circ$ for the setup in Figure 3.25(b). The results are normalized for a unit maximum magnitude (Thiergart, 2007).

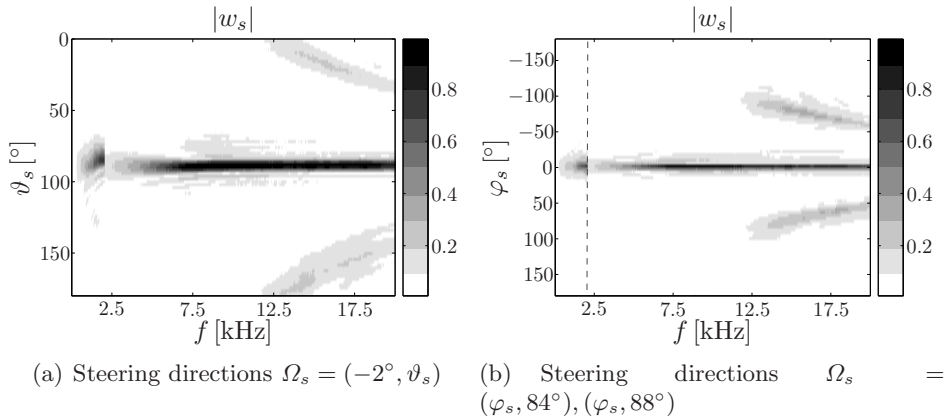


Figure 3.29: Equalized direction weights $|w_s|$ for the vertical and horizontal plane between source and microphone array based on the measurement. Notice the different angular ranges. The vertical offset in the measurement set-up between both microphones is adjusted in (b) (Thiergart, 2007).

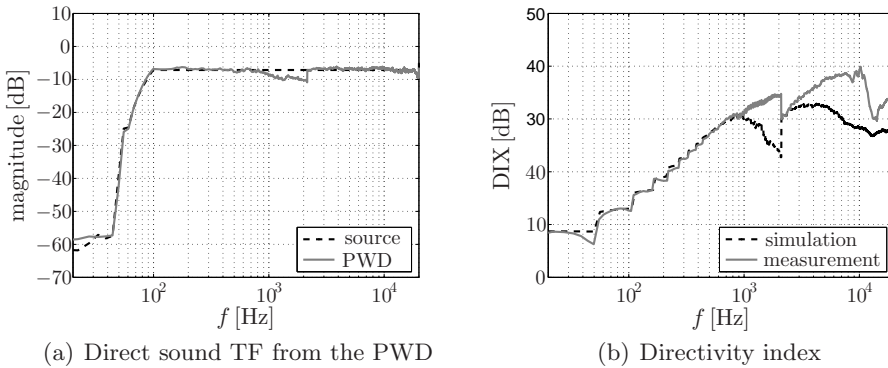


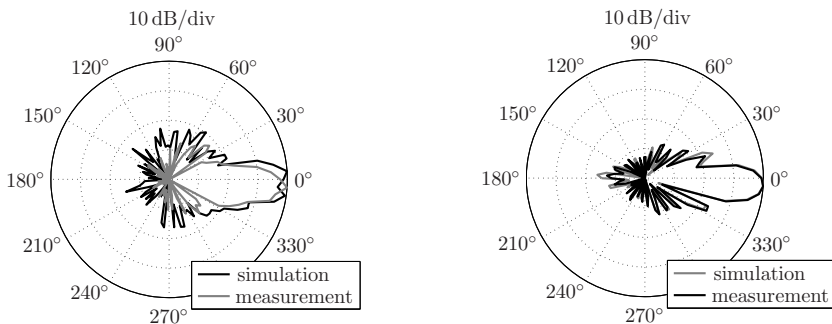
Figure 3.30: (a) Equalized PWD result for the direct sound of the measurement using the setup in Figure 3.25(b) including spatial filtering. (b) The DIX for the same setup (Thiergart, 2007).

pictured separately in Figure 3.30(a), compared to the original spectrum of the source. The transfer function is compensated for a varying maximum level N by applying a spatial low-pass filter. The width of the spatial low-pass filter agrees with the resolution of the PWD, which is expected from theory. The method is explained in detail in Section 4.5.2. The desired unit spectrum of the equalized loudspeaker is nearly obtained. However, it can be noticed that the magnitude decays below the transition frequency $f = 2.1$ kHz. The reason is the unexpectedly low PWD resolution for the radius r_2 . This leads to a wider peak within the direction weights w_s , as visible in Figure 3.27(a). Therefore, a certain amount of the direct sound energy is not included when applying the spatial low-pass filter. This problem is addressed in Section 4.5.2. A general evaluation of the PWD performance can be determined from the directivity index (DIX), depicted in Figure 3.30(b). The DIX of the non-ideal simulation is also presented for comparison. Below approximately $f = 1$ kHz, the DIX is similar for both situations. For $1 \text{ kHz} < f < 2.1 \text{ kHz}$, the DIX of the simulation decays. The reason is a too pessimistic simulation, which significantly reduces the performance of the array at higher kr . For the same reason, the DIX of the measurement is clearly higher compared to the simulation above the transition frequency. However, it decays above $f = 10$ kHz due to spatial aliasing. The highest DIX in practice is 40 dB, in contrast to 45 dB for the ideal simulation and 32 dB for the non-ideal simulation. Furthermore, it can be observed that the DIX for the measurement for r_2 is higher than for r_1 at the transition frequency. This is in contrast to the non-ideal simulation. Hence, the larger

radius r_2 provides a sufficient PWD performance even for the highest operating frequency on this radius. Therefore, the use of the larger radius to improve the resolution at low frequencies is appropriate for practical realizations. The directional responses for the radii r_2 and r_1 at the transition frequency are given in Figure 3.31, compared to the responses for the non-ideal simulation. The main lobes of the responses represent the spatial resolution of the PWD. The measurement results for r_1 in (b) provide a resolution, which is expected from theory. Moreover, the results are almost identical to the simulation. The results for r_2 in (a) display a strong similarity between the measurement and the simulation. The plots in (a) confirm the reduced resolution of the PWD on r_2 for the transition frequency. Figure 3.31 generally shows that the realistic simulation algorithm is able to reproduce the practical measurement effects.

3.8 Summary

This chapter examined the analysis of acoustic fields using spherical arrays. First a short summary of the available beam forming and spatial sampling methods was given. The advantages and disadvantages are considered with respect to the applications required in this work. As a result, a setup using a cardioid open-sphere design was chosen for further investigation. The theoretical basis was given in terms of spherical harmonic decomposition and plane wave decomposition. The errors of practical measurements were inves-



(a) Radius $r_2 = 1$ m, $N = 38$

(b) Radius $r_1 = 0.28$ m, $N = 12$

Figure 3.31: Directional responses $20 \log_{10}(|w_s|)$ at $f = 2.1$ kHz for both array radii, obtained from the non-ideal simulation and the measurement (Thiergart, 2007).

tigated. A taxonomy of measurement arrays was given and their influence on the PWD studied. This analysis includes positioning errors, noise, errors caused by microphone directivity, and spatial sampling errors. It was found that the theoretical frequency-independent resolution of a PWD cannot be achieved in non-ideal (realistic) situations. The robustness of the PWD against measurement problems depends strongly on the maximum level of the spherical harmonic coefficients used to compute the PWD and on the product of wavenumber and array radius kr . The highest robustness against microphone noise and positioning inaccuracies is generally achieved when computing the PWD for $N = kr$. Nevertheless, the influence of spatial aliasing and position error increases for increasing values of kr . Comparing the relative error contributions of the different errors, it can be noticed that spatial aliasing becomes the dominant error at high values of kr . If the maximum level is set to $\lceil kr \rceil$ in the calculations, a small N for low frequency or small array radii is obtained. This reduces the resolution of the array. The frequency range is limited by spatial aliasing in the high frequencies and by measurement noise in the low frequencies and on small measurement radii. This is problematic when considering a wide frequency range, which is necessary in auralization applications. Based on these findings, a new method for the extension of the usable bandwidth in spherical arrays was proposed using a dual radii design. Three methods for the combination of the data acquired by the different array radii have been compared. It was found that a spatial resolution adaptation led to the best results for the combination of two measurement radii to a single data set for wave field analysis. The principles given in this chapter in terms of spatio-temporal acoustic field analysis constitute the basis for the subsequent developments. The theoretical derivations have been validated by measurements using a prototype of a virtual spherical microphone array. It has been shown that the actual measurements agree with the simulations. For this reason, in the following section often simulations using a realistic distribution of measurement errors are considered. This approach gives full control over the effect of different error when evaluating the perceptual consequences. Furthermore, it delivers a controlled environment for evaluation of spherical array systems.

Chapter 4

Room Impulse Response Analysis

4.1 Introduction

This chapter uses the principles derived in Chapter 3 to analyze room impulse response measurements of spherical arrays. It corresponds to block (1) in Figure 4.1 and delivers the source data for block (2). These are the data used for the auralization and interaction with the user. The use of arrays for the analysis of impulse responses has been investigated by other authors with a focus on measurement and visualization to analyze the quality of a reproduction room. Gover et al. (2004, 2002, 2003) proposed an array system for the measurement of directional properties of reverberant sound fields and demonstrated the system in several rooms. Guillaume and Grenier (2006b) compare different two-dimensional array geometries and propose a multi-radii logarithmic array analyzed in (Guillaume and Grenier, 2006a). Roper and Collins (2007) use a combination of a linear and a circular array to analyze the sound field of a room and identify reflections. Rafaely et al. (2007a) have investigated the distribution of reflections in an auditorium using a dual radius open array with omnidirectional sensors. A few authors have discussed the topic of direction-dependent room impulse response analysis for auralization purposes, especially in terms of wave field synthesis (WFS) reproduction. Kuntz (2009) has described the analysis of sound fields using circular arrays. Hulsebos (2004) described in detail the principles based on circular array measurements for WFS applications. His work is based on the previous work and ideas of Sonke (2000) where the separate processing of early reflections and late reverberation has been proposed in general. In the current work, spherical array measurements are used for auralization purposes. The results of the analysis discussed in this chapter should be used for auralization in variable reproduction systems (Chapter 7) as well as for the interactive modification (Chapter 6).

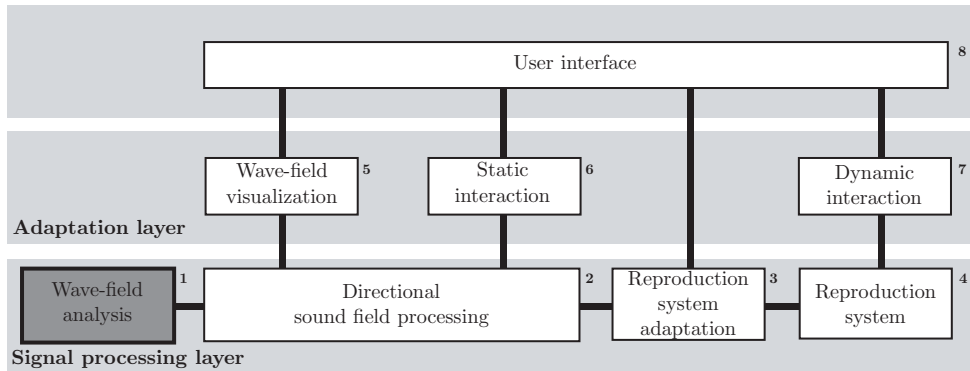


Figure 4.1: Basic block diagram of the system developed in this thesis. The part discussed in this chapter is emphasized.

4.1.1 Outline of the Chapter

First, the required steps of impulse response analysis are introduced in Section 4.2 and discussed in more detail in the subsequent sections. Section 4.3 presents the quality criteria for room impulse response measurements as well as the concept of using a directional speaker for the stimulus during the measurements. The temporal segmentation of the measured room impulse response (RIR) is covered in Section 4.4. Methods to extract strong reflections using dual radius cardioid open spherical array measurements are presented in Section 4.5. Principles for de-noising the diffuse part of impulse responses are given in Section 4.6. The chapter is concluded with archiving the measurements using a new data format based on XML in Section 4.7.

4.2 Basic Processing Chain

Figure 4.2 presents a block diagram of the RIR analysis process. The measurement data are acquired using a spherical microphone array as described in the previous chapter. The basic steps are:

1. **Measurement and system equalization:** The system equalization is applied to remove the influence of the measurement chain (e. g., loud-speaker, microphone).

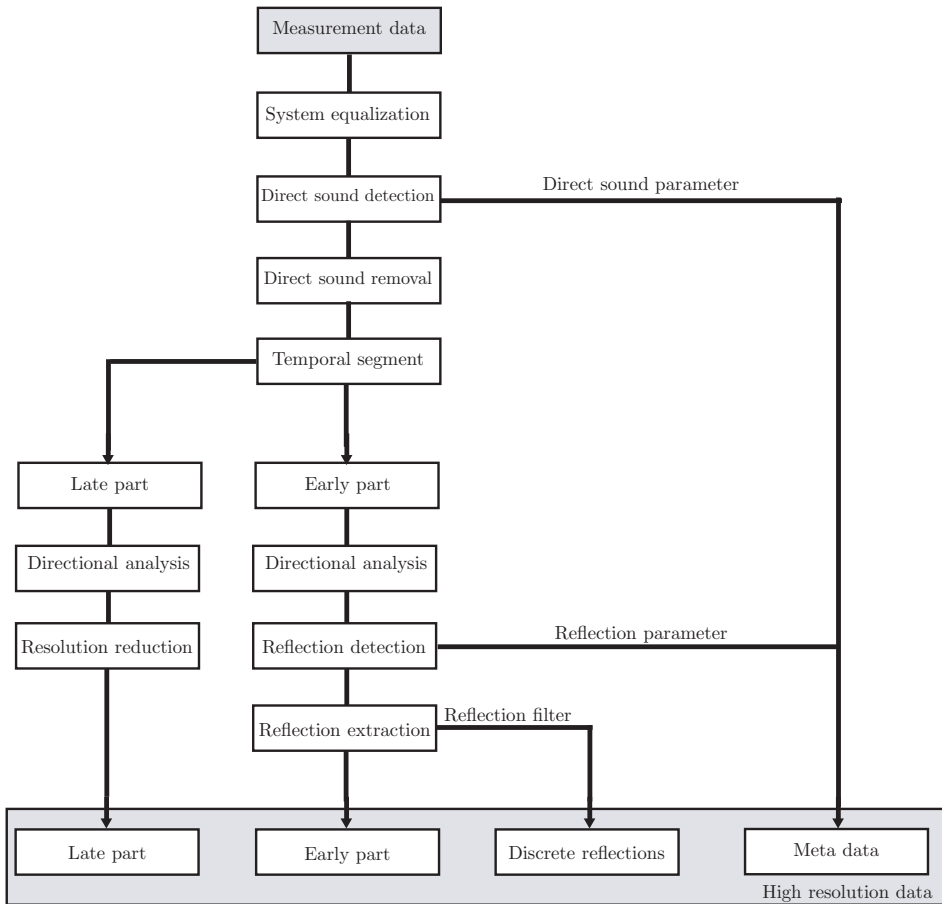


Figure 4.2: Block diagram of the impulse response analysis process.

2. **Detection and removal of direct sound:** The direct sound is removed by filtering in the time domain. This part will be recreated during the auralization process. The parameters of the direct sound are stored as meta data for future reference.
3. **Segmentation:** To optimize the computation, the impulse responses are split into two time segments, the early part and the late part.
4. **Analyzing the early part:** A directional analysis is applied to the early part. This corresponds to a plane wave decomposition as discussed in Section 3.4. The resolution can be adapted to the needs of the aural-

ization and interaction process as will be discussed in the next chapters.

5. **Detection and extraction of reflections from the early part:** The early part of the RIR is used to extract important discrete reflections. These are stored separately.
6. **Analyzing the diffuse part:** A directional analysis is applied to the late part. This corresponds to a plane wave decomposition as given in Section 3.4. The resolution can be adapted to the needs of the auralization and interaction process as discussed in the next chapters.

These steps will be discussed in the next sections and the application of the theoretical principles given in the previous chapters will be shown. The high resolution data¹ is archived after the analysis process. In this data set, all filter and corresponding meta data are stored in an XML-file while separate audio files contain the impulse responses.

4.3 Measurement and System Equalization

When measuring impulse responses for auralization purposes, the most important quality criteria for measurements can be summarized as follows:

- The *signal-to-noise ratio* should be as high as possible.
- The *frequency response* of the measurement system should be flat and broad.
- A good *repeatability* is desired.

An overview of different measurement signals and their advantages and disadvantages can be found in (Müller and Massarani, 2001). By using logarithmic sweeps as proposed by Farina and Trochlin (2004), the desired quality criteria can be achieved. Hulsebos (2004) has shown that these measurement techniques are an adequate solution for room impulse measurements to be used for auralization. If measurements are made using virtual arrays, it is important to keep the measurement time per single impulse response relatively short because a huge number of impulse responses has to be acquired. For this reason, a high sound pressure level in the reproduction is required to achieve a sufficient signal-to-noise ratio (SNR). A high sound pressure level in

¹The data set is named *high resolution* because of its high spatial resolution.

a broad frequency range is difficult to achieve with available omni directional dodecaeder loudspeakers, or requires strong equalization (Farina and Ayalon, 2003). An alternative is to use public address systems² or large monitor loudspeakers (Kessler, 2005). Thesis loudspeaker systems are not omnidirectional in the higher frequency range but have controlled directivity. For this reason, different measurements with a rotated loudspeaker at one source position have to be combined. The post-processing steps are the same as for a single measurement. The combination will be discussed in the next paragraph. If the impulse response is used for auralization, only the transfer function of the room should be included in the data. For this reason, the measured impulse responses have to be equalized to remove the frequency response of the measurement equipment (Müller and Massarani, 2001). Ideally one array measurement is made for each required source position. Due to the time-consuming measurement process using spherical arrays, methods to overcome a large number of different measurements will be discussed in Section 6.4.4.

Combination of directional speaker measurements Directional speakers can offer a flat on-axis frequency response and a high sound pressure level. In case of a cardioid directivity, the measurement has to be performed four times to excite all reflections, while the loudspeaker is rotated for every measurement by 90° . The high resolution data of all four measurements are combined by superposition to achieve a complete set of measurements for one source position. An example of two loudspeaker directions is given in Figure 4.3. The strong excitation of a wall reflection can be identified clearly in plot (b). The combination of the measurements can be performed in two ways:

1. The measurements are weighted according to the desired frequency dependent directivity of the virtual source that will be reproduced³.
2. Only a limited number of discrete reflections are extracted and varied according to the simulated source directivity. Additionally, these reflections can be used in a pseudo-extrapolation process as given in Section 6.4.4. To enable this approach, the strongest first-order discrete reflections are extracted from each measurement. It is taken into account that the reflections in direction of the main axis of the speaker are excited very strongly due to its directivity. Each measurement is used to extract a subset of reflections. The reflections are removed as described in Section 4.5. Since

²Speaker systems designed for sound reinforcement.

³This approach was first proposed to the author by Dietz Tinhof.

the directivity of the speaker is not ideal, the corresponding reflection positions have to be removed in the other measurements as well.

As pilot experiments (see Chapter 8) have shown, the correct reproduction of the first-order reflections is sufficient to simulate different source positions and directivity in a high perceptual quality.

4.4 Segmentation

In this processing step, the transition time between the domain of early reflections and the diffuse part of the impulse response has to be determined. In the literature, the value of 50 ms and 80 ms is often given as the *transition time* or *mixing time* (e. g., (Begault, 1992)). Meesawat and Hammershoi (2003) have investigated the perceptual threshold for the late part in a listening experiment. The design was a three alternative forced choice experiment. The task for the subject was to identify any perceptual difference in the three alternatives. The result of the tests, using measurement taken in a small lecture room, was 40 ms. This result will now be compared with criteria proposed in the literature. Several authors have proposed a criterion based on the *reflection density*. The number of reflections can be calculated for a rectangle mirror source model by (Möser, 2007):

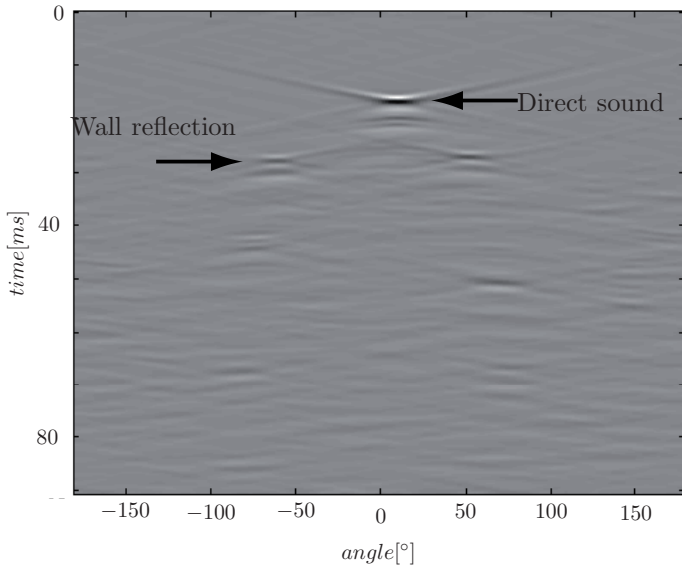
$$\frac{\Delta N}{\Delta t} \approx \frac{dN}{dt} = 4\pi c \frac{(ct)^2}{V}, \quad (4.1)$$

where N denotes the number of reflections, c is the speed of sound, V the room volume and t is the time. A second criterion used for the prediction of the transition time between the early and late or diffuse part of the impulse response is the so-called *mean free path* as given by

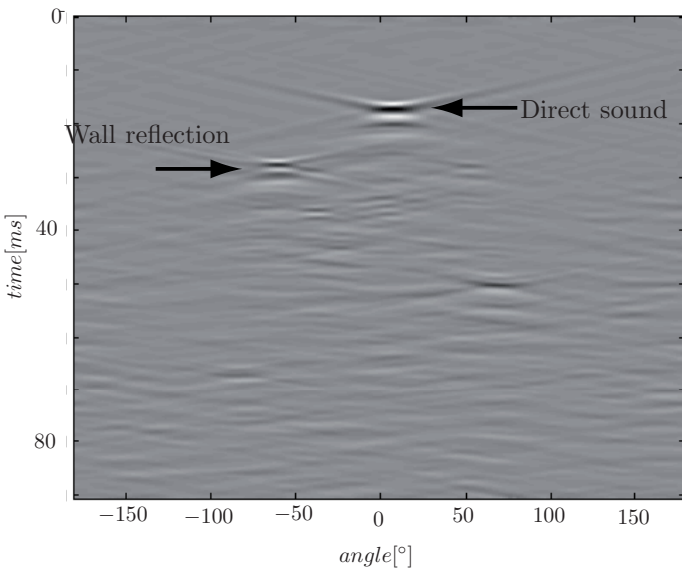
$$l_m = 4 \frac{V}{S}, \quad (4.2)$$

where S represents the total boundary surface area. Eq. (4.2) is valid in a room with somewhat diffusely reflecting walls as stated by (Kuttruff, 1991) as well as for flat walls (Kosten, 1960). Based on this measure, the following criteria have been compared with the results of the listening experiments by Meesawat and Hammershoi (2003):

1. A transition time based on the the reflection density was given by Schroeder (1962). He used 1000 reflections per second as threshold.



(a) Main axis of the loudspeaker pointing towards the array.



(b) Main axis of the loudspeaker pointing toward the side wall.

Figure 4.3: Horizontal plane wave decomposition of a sound field in a small lecture room when using a directional speaker.

2. Griesinger (1989) proposed a density of 10000 reflections per second.
3. Rubak and Johansen (1998) proposed a criterion of 4400 reflections per second.
4. Rubak and Johansen (1999) proposed a time criterion of four times the mean free path.

Compared to the results of the listening experiment of Meesawat et. al. (2003), the mean free path criterion (4) as well as the 4400 reflection criterion (3) predict the transition time correctly. Knowledge of the room geometry is required for all of these criteria in order to calculate the reflection density or the mean free path. In this work, a solution independent of geometry is desired. Furthermore, the results have been proven only for a small lecture room and not for different rooms as a more general solution. For these reasons, a less stringent criterion is desired.

A geometry-independent solution was proposed by Ben-Hador (2004). They use the time RT_{10} over which the Schroeder integral as given by Eq. (6.3) has decayed by 10 dB. The transition time is then defined as: $\min[RT_{10}, 80 \text{ ms}]$.

Abel and Huang (2006) proposed an approach based on the assumption that the sound pressure amplitudes in a reverberant field take on a Gaussian distribution. A so-called ‘echo density profile’ is calculated for determining the mixing time: With a short sliding rectangular window of 500-2000 samples, the empirical standard deviation of the sound pressure amplitudes is calculated for each sample. In order to determine how well the empirical amplitude distribution approximates a Gaussian behavior, the proportion of samples outside the empirical standard deviation is determined and compared to the proportion which is expected for a Gaussian distribution. This echo density profile should increase with increasing time and diffusion. At the instant of complete diffusion, it reaches the value of one. The gross shape of the echo density profile stays similar with larger window sizes and some smoothing can be observed. A rectangular shape window of 210 samples (23 ms) was suggested by Abel and Huang (2006) from discussion of auditory time resolution. The mixing time $t_{mix,Abel}$ can be defined as the instant where the echo density profile becomes unity for the first time.

More recently Lindau et al. (2010) evaluated the correlation of physical predictors and perception of mixing time in binaural reproduction. Lindau et al. (2010) gives a good overview of empirical predictors of physical mixing time. They found by a perceptual experiment which was performed for several room

sizes of shoe box shaped rooms two highly correlated formulas. The following empirical equation can be used if the geometry of the room is known,

$$t_{mix} = 0.0117 \cdot V + 50.1, \quad (4.3)$$

where the t_{mix} is the mixing time in milliseconds. In case the impulse response is available, a formula based on the findings of Abel and Huang (2006) results in the regression formula:

$$t_{mix} = 1.8t_{mix(Abel)} - 38 \quad (4.4)$$

The relative reverberance was not found to have a significant influence. In this work, the approach of Lindau et al. (2010) is used to segment the impulse responses because the required data can be calculated directly from the impulse response. For the transition between the two parts of the impulse response, a cross-fade is required.

4.5 Detection and Extraction of Reflections

The early part of the room impulse response is used to detect and extract discrete reflections. This process will be studied using a simulation based on the geometry in Figure 4.4(a). Only a direct sound source and a single wall reflection are modeled in an anechoic environment. Both are assumed to be point sources. The direct source emits an ideal white spectrum, while the reflection results in a high-pass filtered spectrum. The wall indicated in Figure 4.4(a) has a limited width. Therefore, it acts as a spectral high-pass filter. The intensity of the reflected sound is reduced at low frequencies due to diffraction effects (Dickreiter, 2008, p. 16). The lowest reflected frequency can be estimated using (Dickreiter, 2008; Meyer, 1995, p. 16)

$$f_{\min} = \frac{2c}{(d \cdot \cos \beta)^2} \cdot \frac{a_1 \cdot a_2}{a_1 + a_2}, \quad (4.5)$$

where c denotes the speed of sound, $d = 2.45$ m the width of the wall, and $\beta = 22.5^\circ$ the incident angle of the sound. The distances between the wall and the loudspeaker and the wall and the listener are denoted by the symbols $a_1 = 6.67$ m and $a_2 = 6.75$ m, respectively. In the simulation set-up in Figure 4.4(a) $f_{\min} = 450$ Hz is obtained. For all frequencies $f > f_{\min}$, it is assumed that the reflected sound exhibits the same spectrum as the source.

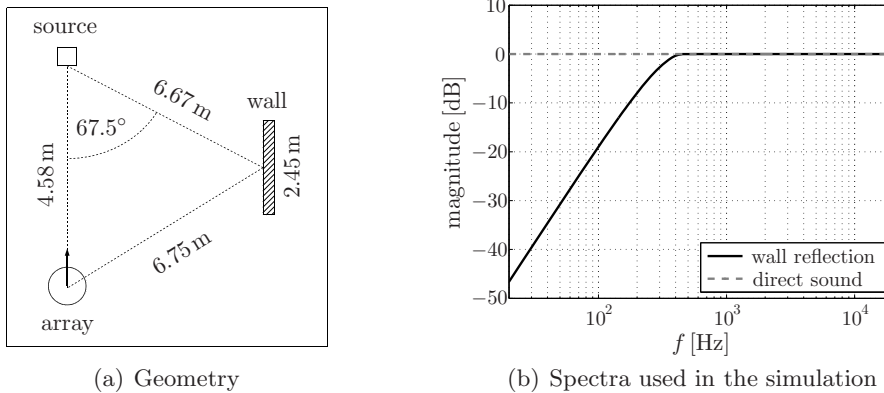


Figure 4.4: (a) Top view of the simulation setups. The arrows denote the azimuth angle $\varphi_s = 0$. (b) Transfer functions of the direct sound and the wall reflection. Only the positive frequency part of the spectrum is plotted (Thiergart, 2007).

The transfer function (TF) of the reflected sound, which is considered in the following section, is presented in Figure 4.4(b). The simulation uses a cardioid open sphere microphone array with 2 radii ($r_1 = 0.28$ m, $r_2 = 1$ m) on a Lebedev grid. $Q = 2030$ positions are used in a frequency range up to 20 kHz. The aliasing frequencies of the different radii are $f_1 = 7.4$ kHz and $f_2 = 2.1$ kHz, respectively. The transition frequency was set to f_2 . Two situations are studied, an ideal situation and a non-ideal situation including the following errors:

- Random azimuthal position error with a standard deviation of 1°
- Random elevation position error with a standard deviation of 4°
- Offset error in elevation and azimuth of 3°
- Normally distributed white noise with 80 dB SNR on the microphone signals
- Non-ideal cardioid characteristic according to (Sennheiser, 2007c)

The impulse responses are generated by superposition of the direct and reflected source field. Air damping were not included.

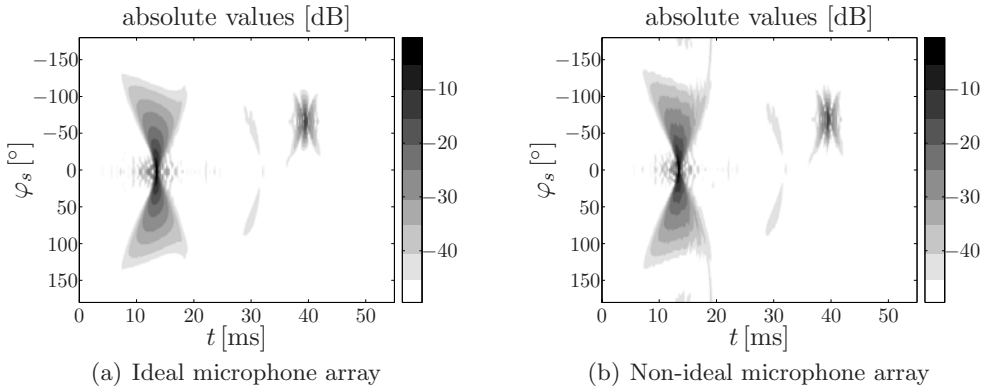


Figure 4.5: Impulse responses for the horizontal plane between source, wall, and microphone array (simulation) (Thiergart, 2007).

4.5.1 Detection of Reflections

Figure 4.5 presents the magnitude values of the PWD. This representation can be used to detect the direction of arrival (DOA) and time of arrival (TOA) of the reflections. In a non-ideal situation, the analysis has to be performed for each elevation angle by identifying the peaks in the PWD. Furthermore, the near-field effects of the point source simulation can be noticed. While the energy of the reflection is very concentrated to a point, the energy of the direct sound is spread over a wider angular range. This is caused by the fact that a point source at a distance of 4.58 m was simulated. The reflection on the other hand is further away and resembles a plane wave when arriving at the array.

4.5.2 Extraction of Reflections

The results of the detection process are the DOAs and TOAs of the reflections in the early part of the room impulse response. Once the positions of reflections are detected, they should be extracted and removed to be individually stored and processed. To characterize the frequency response of the reflection, a filter is required. In this section, the extraction of the filter is studied in detail using simulations and measurements. First the extraction of a pure direct sound is considered. Subsequently, a scenario with direct sound and a single reflection is analyzed. The results of these simulation-based scenarios are verified with a measurement. To extract a reflection a time window is used

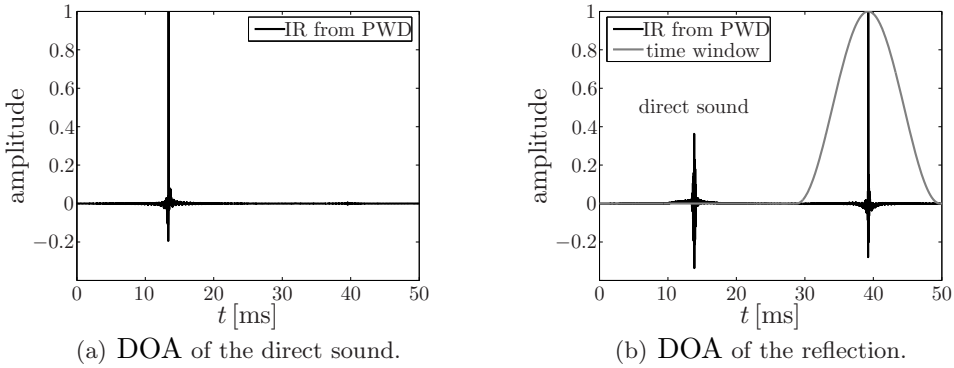


Figure 4.6: Single impulse responses obtained from the direction weights in the case of an ideal microphone array. For temporal separation the responses are filtered by applying a Hanning window (Thiergart, 2007).

first for the temporal separation. Figure 4.6 shows the use of a Hann window on the reflection. For illustration purposes only, the impulse response in DOA of the reflection is shown with a 21 ms Hann window. Before a filter for the reflection can be extracted, the spatial resolution has to be taken into account. First, only the direct sound is considered. Figure 4.7 shows the simulation without reflections. The magnitude spectra of the directional weights (PWD)

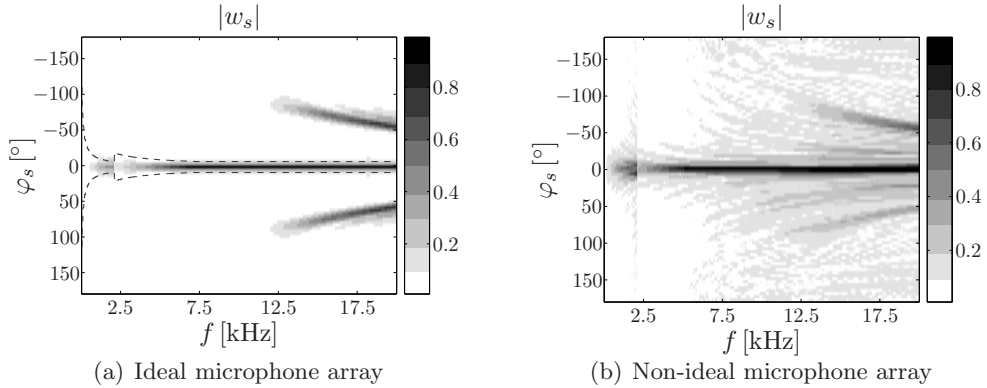


Figure 4.7: Direction weights $|w_s|$ as a function of f for a co-elevation $\vartheta_s = 90^\circ$ for the setup in Figure 4.4(a) considering the direct sound only. The results are normalized to maximum magnitude (Thiergart, 2007).

are given for the horizontal plane of the source. Discontinuities, as discussed

in Section 3.6, can be noticed at the transition frequencies between the array radii. Furthermore, strong aliasing can be noticed in the ideal situation. The aliasing error is problematic if additional reflections reach the array from other directions. Interferences will arise between the transfer function of the reflection and the aliasing error. In the non-ideal simulation, a significant amount of error can be noticed, even in lower frequencies. Since the frequency-dependent level in the calculation of the PWD was set to $N = \lceil kr \rceil$, the non-ideal results prove a high array robustness against microphone noise and positioning errors as derived in Chapter 3. In Figure 4.8(a) the used values for the level N are given. The resulting theoretical values for the resolution are shown in Figure 4.8(b) (see Section 3.4.1). The spatial width (spatial resolution) of the PWD is also shown in Figure 4.7, marked with a dashed line. A simple

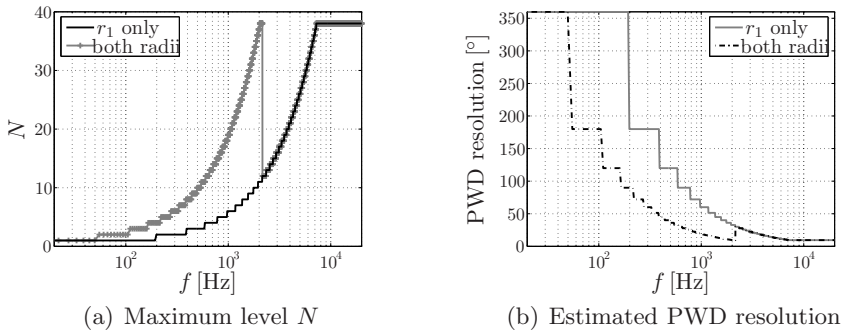


Figure 4.8: Plot (a) shows the maximum level used in the calculations as a function of f for the cases of using only r_1 or a combination of two radii. The PWD resolution, which is expected from (a), is depicted in (b) (Thiergart, 2007).

method to extract the frequency response of a reflection is the separation of the directional weights corresponding to the DOA. This will result in the response seen in Figure 4.9(a). The transfer function features a high-pass characteristic. This is due to the frequency-dependent resolution, which causes energy to be spread over a broader range of directional weights and not only the DOA of the reflection. The original spectrum is obtained if the direction weights of all directions are summed (see Figure 4.9(b)). Since spatially separated events occur simultaneously, this is not possible in real applications. Furthermore, the non-ideal simulation shows significant discrepancies at frequencies with high kr , which lead to a lower SNR. A better approach to separate spatially separated events is to apply a spatial low-pass filter. To keep the main section of

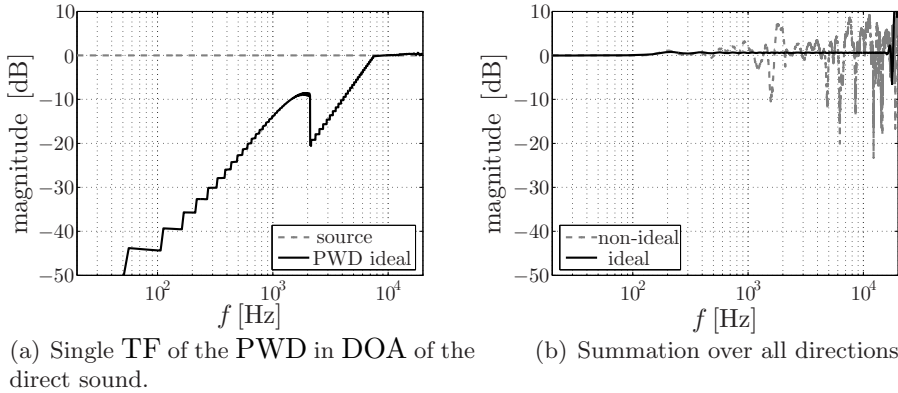


Figure 4.9: (a) TF of the direct sound obtained from the PWD and the original TF of the source. (b) PWD results for the ideal and non-ideal array after integrating over all direction weights (Thiergart, 2007).

the energy, the range for the summation should be equivalent to the resolution of the PWD. The spatial low-pass filter is three-dimensionally extended in all directions. The required width of the filter can be approximated using Eq. (3.45). The result is a frequency-dependent spatial low-pass filter with a width $\Theta(kr)$ given by

$$\Theta(kr) \approx 2 \frac{\pi}{N(kr)}, \quad (4.6)$$

where N is the frequency-dependent level used in the calculations of the PWD. In the examples it is set to $N(kr) = \lceil kr \rceil$. The result of such a filter is given in Figure 4.10(a). It shows that filtering results in a correct magnitude response. Furthermore, the transfer function decays below the transition frequency because the resolution on r_2 is smaller than expected at higher frequencies. Therefore, a substantial amount of energy is located outside the filtered area. The non-ideal situation is shown in Figure 4.10(b). Beside strong similarities, the magnitude values also decay above the transition frequency and begin to rise slightly at high frequencies. The reason for the latter behavior is the inclusion of aliasing energy in the spatial low-pass filter. To optimize the results, the width of the spatial low-pass filter can be increased with $\eta(kr)$ resulting in

$$\Theta(kr) \approx 2 \frac{\pi}{N(kr)} + \eta(kr). \quad (4.7)$$

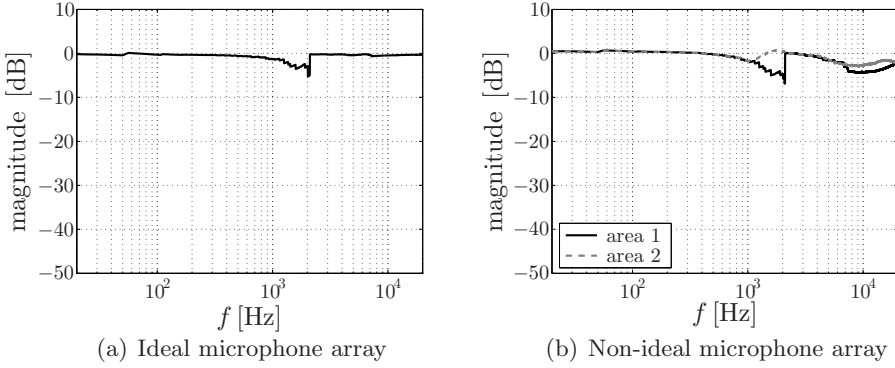


Figure 4.10: Transfer function extracted from the PWD results for $\Omega_s = (0^\circ, 90^\circ)$ after applying two different spatial low-pass filters over neighboring direction weights (Thiergart, 2007).

A comparison is shown in Figure 4.10(b). For the frequency response denoted as area 1 the following expression is used:

$$\Theta(kr) = \begin{cases} \frac{2\pi}{N(kr)} & \forall \Theta \geq 10^\circ \\ 10 & \forall \Theta < 10^\circ. \end{cases} \quad (4.8)$$

In the second case denoted as area 2, the minimal resolution of the filter was set to 15° instead of 10° by using

$$\Theta(kr) = \begin{cases} \frac{2\pi}{N(kr)} & \forall \Theta \geq 15^\circ \\ 15 & \forall \Theta < 15^\circ. \end{cases} \quad (4.9)$$

With this correction term the frequency response is optimized. The result may be less convincing if there are interfering events from other directions, aliasing, or noise. Next, the original simulation, including the reflection, is considered again. Figure 4.11 presents the normalized directional weights. The results after applying a spatial low pass of the ideal shape are presented in Figure 4.13 for the direct sound and in Figure 4.12 for the reflections. The last step is the subtraction of the extracted direction weights from the early part of the impulse response. The extracted TF of the reflections are stored with the DOA in the discrete reflection section of the high resolution data. The extraction process can be repeated until the desired number of reflections is reached.

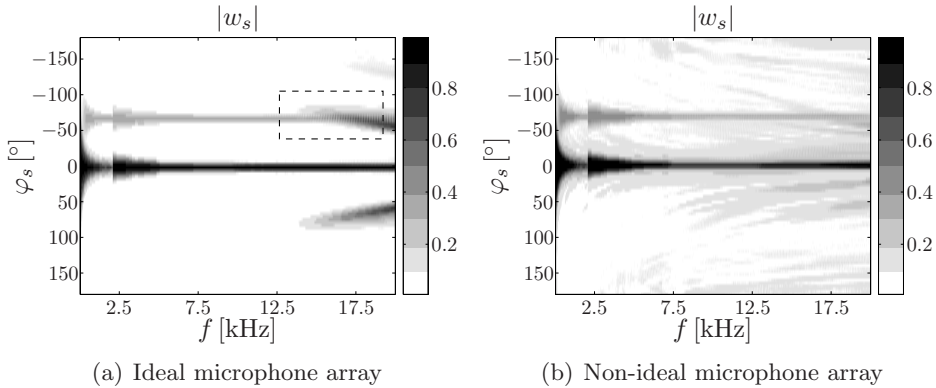


Figure 4.11: Normalized directional weights $|w_s|$ as a function of f for a co-elevation $\vartheta_s = 90^\circ$ for the simulation setup in Fig. 4.4(a). The energy of the direct sound, arriving from $\varphi_s = 0^\circ$, is spread over a wide area at frequencies for which the PWD resolution is low. Strong interference between the aliasing components of the direct sound and the wall reflection are marked with a rectangle (Thiergart, 2007).

4.5.3 Extraction of Reflections Based on Measurements

The simulations of the previous section have been verified by a measurement under the same conditions as in Section 3.7. The loudspeaker used during the measurements was a Kling&Freitag CA106 (Freitag, 2000). This loudspeaker was used because it is a directional loudspeaker for sound reinforcement, as described in Section 4.3. The CA106 is a 2-way speaker, which is small enough to be measured in the far-field in terms of combination of the tweeter and the woofer. The free-field transfer function is used as a reference to compare the results of the extraction. The transfer function was determined from the impulse response of the array microphone, which points exactly towards the loudspeaker. For this purpose, the setup as given in Section 3.7 was considered. The larger radius r_2 microphone is used to avoid any influences of the robot arm. The results are depicted in Figure 4.14(a). It can be noticed that the original spectrum of the source in Figure 4.14(a) is not perfectly white. For this reason, an equalization has to be applied as described in Section 4.3. It can be performed before computing the spherical harmonic decomposition (SHD) and plane wave decomposition (PWD). To this end, the outputs of all microphone positions are equalized (Müller and Massarani, 2001). The equalization can also be applied to the PWD results in the post-processing. This can reduce

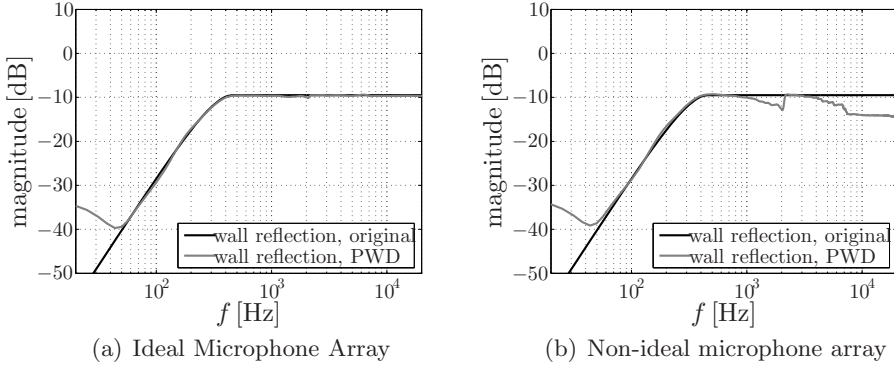


Figure 4.12: TF of the wall reflection, determined from the PWD, including a time domain window (Thiergart, 2007).

the computational effort, for example, if only a few steering directions are extracted. This method is used as follows. To compute the equalization, the PWD results are divided by the magnitude spectrum of the original loudspeaker transfer function in Figure 4.14(a). To avoid a high noise amplification, only the frequency range $100 \text{ Hz} \leq f \leq 20 \text{ kHz}$ is equalized. In this frequency range the loudspeaker transfer function exhibits a high SNR. The equalized transfer function of the loudspeaker is also shown in Figure 4.14(a). As a reference, Figure 4.14(b) shows the transfer functions of the microphone positions $\Omega_q = (0^\circ, 90^\circ)$ and $\Omega_q = (-67.5^\circ, 90^\circ)$. They represent the transfer functions of the direct sound and wall reflection. In addition to the equalization, a Hann window (Harris, 1978) of width $T = 18 \text{ ms}$ is applied. As shown in Figure 4.14(b), above approximately $f = 1.5 \text{ kHz}$ a decay of the direct sound takes place as compared to the wall reflection. This is a result of the directivity of the loudspeaker which was directed towards the wall. Furthermore, the high-pass characteristic of the wall is visible below $f \approx 450 \text{ Hz}$, where the magnitude of the wall reflection decreases in contrast to the direct sound. Figure 4.15 shows the direction weights w_s at the transition frequency $f = 2.1 \text{ kHz}$ for both array radii, demonstrating the highest and lowest PWD resolution for r_2 and r_1 , respectively. The direct sound and the wall reflection can be separated in both cases. For the radius r_2 , the direct sound and the wall reflection are obtained for the directions $\Omega_s^{(d)} \approx (-2^\circ, 86^\circ)$ and $\Omega_s^{(w)} \approx (-67^\circ, 86^\circ)$. Apart from slight positioning errors, these directions agree with the setup in Figure 4.4(a). For the radius r_1 , the direct sound and the wall reflection arrive at the array at $\Omega_s^{(d)} \approx (-2^\circ, 88^\circ)$ and $\Omega_s^{(w)} \approx (-67^\circ, 88^\circ)$. Hence, both microphones possess

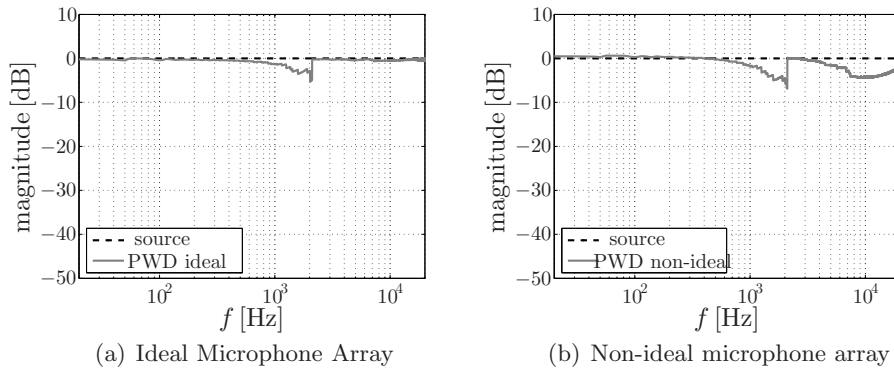


Figure 4.13: TF for the direct sound ($\Omega_s = (0^\circ, 90^\circ)$) extracted from the PWD, in contrast to the original TF of the source. The extraction was performed using compensation for a varying N by applying a spatial low-pass over neighboring directional weights (Thiergart, 2007).

a vertical misalignment of about 2° . The value of the error is smaller than in the previous section, since the microphones were readjusted before starting the measurement. Next, the extraction of the TF of the reflection and the direct sound is studied. Figure 4.16 shows the equalized PWD output for the horizontal plane between source, wall, and microphone array. The vertical offset is compensated by using the steering directions $\Omega_s = (\varphi_s, 86^\circ)$ below the transition frequency and the directions $\Omega_s = (\varphi_s, 88^\circ)$ above the transition frequency. The transition frequency is marked by the dotted line. The transfer function of the direct sound and the wall reflection can be observed clearly. The impulse responses, which correspond to the transfer functions in Figure 4.16(a), are illustrated in Figure 4.16(b). As expected from the geometry, the impulse response of the wall reflection arrives at the array with a delay of $\tau \approx 25$ ms after the direct sound. It can be noticed further that the energy of the direct sound impulse response is spread over the DOA of the wall reflection. The transfer function of the wall reflection is depicted in Figure 4.16(a) for $\varphi_s = -67^\circ$. A spatial-low pass filter has been applied beforehand. The width of the low-pass filter agrees with the theoretical resolution of the PWD as given in Eq. (4.6). Figure 4.17(b) shows the reference signal in contrast to an extracted version. This was spatially filtered and windowed in the time domain using a Hann window of width $T = 18$ ms. A strong similarity is seen. The results generally show that it is possible to reconstruct the original transfer function of the sound, using the PWD. Next the direct sound is examined. A spatial-

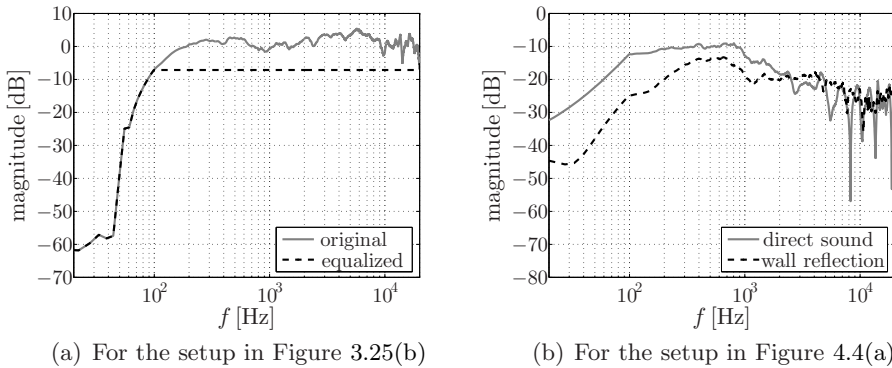


Figure 4.14: (a): Transfer function of the Kling & Freitag CA 106. (b): Transfer functions for the microphone positions pointing towards source and wall. All transfer functions are windowed in the time domain and equalized in the frequency domain (Thiergart, 2007).

low pass and a Hann window of width $T = 18$ ms was applied. The results are depicted in comparison with the reference signal in Figure 4.17(a). The PWD results show a strong similarity with the reference signal. Below the transition frequency $f = 2.1$ kHz, a small deterioration of the PWD results is observed. The reason is the unexpectedly low resolution of the PWD for the direct sound on r_2 , as visible in Figure 4.15(a). Thus, a specific amount of direct sound energy is not considered when computing the spatial low-pass filter. Nevertheless, the results show that the PWD is an appropriate approach for reconstructing the original transfer functions of the arriving sound.

4.6 De-noising of the Diffuse Part

The signal-to-noise ratio of the measured room impulse response is critical for the diffuse or late part of the impulse response. Towards the end of the late part the diffuse energy drops into the noise floor of the measurement. For de-noising two methods are given. One possible solution is to fade out the impulse responses in the time domain before the noise floor is reached. A disadvantage of this approach is that due to the frequency dependent decay time some noise is left in the impulse response, especially in the high frequency range where the reverberation time becomes shorter due to air absorption. An optimization proposed by Meesawat and Hammershoi (2003) is a frequency-dependent

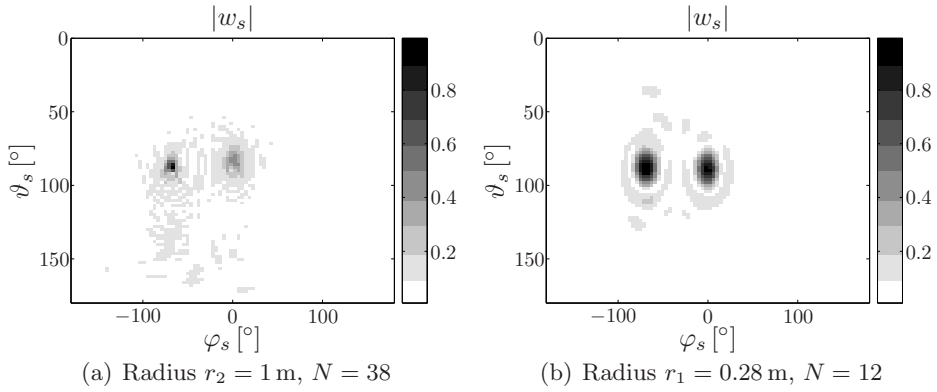


Figure 4.15: Normalized direction weights $|w_s|$ at a frequency $f = 2.1$ kHz for the measurement setup in Figure 4.4(a). The left peak in each plot represents the wall reflection (Thiergart, 2007).

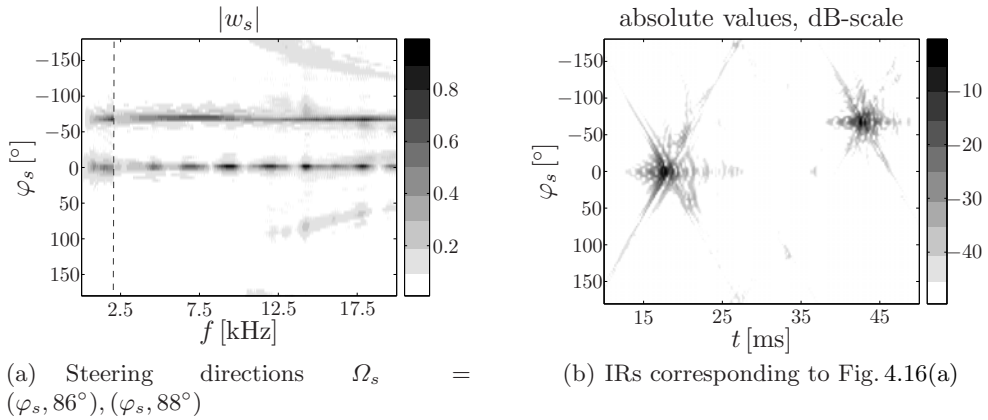


Figure 4.16: Plot (a): Equalized directional weights w_s for the horizontal plane between wall, source, and array. The microphone misalignments are corrected. Plot (b): Corresponding impulse responses (Thiergart, 2007).

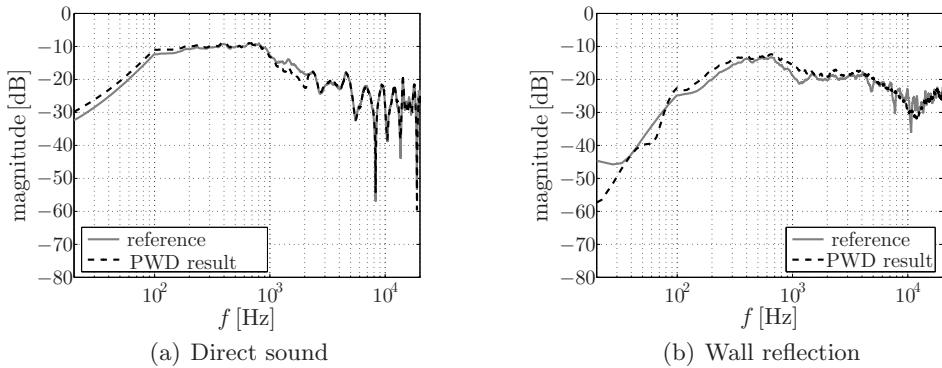


Figure 4.17: Reference and transfer function determined from the PWD for the DOA of the direct sound (a) and the wall reflection (b) using spatial and temporal filtering (Thiergart, 2007).

fading of the noise floor. While the reverberation time is calculated from the early diffuse part, the late part including the noise floor is faded dependent on the frequency in order to preserve the decay time for each frequency. This operation can be performed in the short time Fourier transform (STFT) domain (see Section 6.3.2). Another approach was proposed by Jot et al. (1997). They suggested de-noising by extension of the reverberation tail with shaped noise signals. After the transition time between the reverberation tail and the noise floor is approximated, the decay will be extended according to the frequency-dependent reverberation time using a shaped noise signal. An advantage of this approach is that the reverberation time can be modeled exactly, until the resolution limits of the processing is reached. Within this work the latter approach was used to de-noising the late part of the impulse response if required.

4.7 Archiving of Array Measurements

A XML-format for the storage of room impulse response measurements and analyzed data was developed. The format named as *high resolution data* is able to store the three different parts of the impulse response. Namely the late part, the early part, and the discrete reflections. The positions of the reflections and the direct sound are also stored. A complete data set is composed of the filter coefficients (RIRs) and meta data required for the processing and reproduction.

The *meta data* are stored in an XML-file and include information about the room and required global geometry for the pseudo-extrapolation as discussed in Section 6.4.4. The filter coefficients are stored in linked audio-files. The different parts of the processed signals are stored in dedicated sections (XML-nodes) in the file. The data format developed is flexible and can be used with different array geometries and reproduction systems. It is operated via the user interface developed in Section 6.5.

4.8 Summary

In this chapter the methods of sound field analysis using spherical microphone arrays were applied to the measurement of room impulse responses. For this application, a post-processing chain of room acoustic measurements using spherical arrays was given in the first section. The processing steps required were studied in detail in the subsequent sections. These consist of measurement methods given by the literature and appropriate quality criteria. A concept for using a directional loudspeaker for room impulse response measurements was described. It was shown that the application of such speaker is useful to acquire direction-dependent room impulse responses for a specific source position. In the further processing such measurements can be used to simulate different directivities as well as a data set for an optimized extraction of single reflections. Methods to calculate the transition time of the early and late part of the impulse responses have been reviewed. A method was chosen with respect to a very practical implementation. The main part of this chapter studied the detection and extraction of strong discrete reflections of spherical array measurement data. Methods for spatio-temporal windowing have been applied to simulations as well as to measurements. It has been shown that the extraction of transfer functions for strong discrete reflections is possible in ideal as well as non-ideal conditions. The simulations have been verified by a measurement under controlled conditions in an anechoic room. Finally, after a discussion of de-noising methods for the late part of the impulse responses, a new data format to store measurement and post-processed high spatial resolution data was introduced. The methods developed in this chapter deliver the basic data for the modification of measurements by a sound designer. This will be discussed in the following chapter. Furthermore, a high resolution data set of room impulse response data can flexibly be adapted to different reproduction systems as studied in Chapter 7.

Chapter 5

Wave Field Extrapolation

5.1 Introduction

In auralization applications it is desirable to reconstruct a measured sound field for arbitrary positions (e. g., for the simulation of virtual microphones (VM) as discussed in detail in Section 7.3). This chapter corresponds to block (2) in Figure 5.1. The approaches studied in this chapter allow the extrapolation

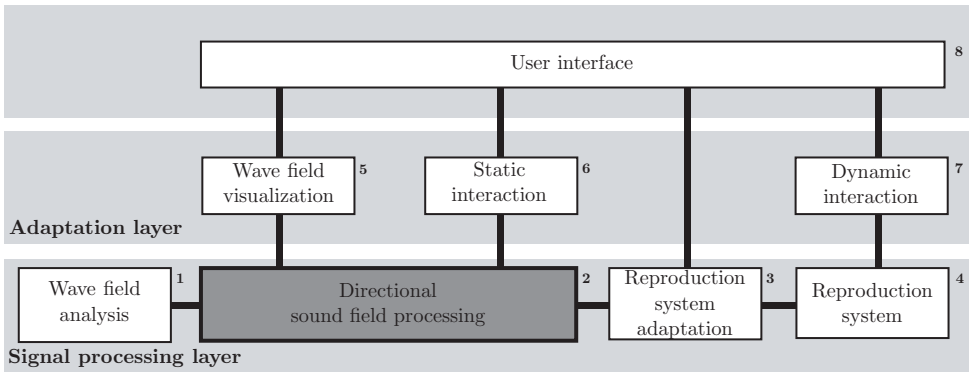


Figure 5.1: Basic block diagram of the system developed in this thesis. The part discussed in this chapter is emphasized.

of sound fields to a desired position. This is done for pressure and particle velocity based on spherical harmonics decomposition. A second approach based on plane wave extrapolation including different directivity functions has also been studied. Based on these methods, the effects of level-limited fields and measurement errors are analyzed.

5.1.1 Outline of the Chapter

Two approaches for wave field extrapolation are analyzed. In Section 5.2 the extrapolation based on plane wave decomposition (PWD) and in Section 5.3 the extrapolation based on spherical harmonics are presented. While the latter is used for the calculation of a VM inside the extrapolated field as described in Section 7.3, the former is important in case of binaural adaptation as discussed in Section 7.4. The approaches are compared in Section 5.4. As discussed in Chapter 3, only a finite resolution can be achieved using microphone array measurements. This results in a limited size of the area of accurate extrapolation (AAE). This, and other possible errors are analyzed in Section 5.5. Beside the AAE limits, the influence of extrapolation on the statistical properties of an acoustic field is relevant for the late part of a room impulse response. For this reason, the influence of aliasing and extrapolation errors on the coherence of an extrapolated field is studied in Section 5.6.

5.2 Plane Wave Extrapolation

The plane wave decomposition results can be used to extrapolate the pressure of a sound field to an arbitrary point (R, φ, ϑ) , the result being denoted as $P(R, \varphi, \vartheta, k)$. Since the direction weights $w_s(\phi, \theta, k)$ represent the magnitude of plane waves with wavenumber k arriving from direction (ϕ, θ) , the sound field can be extrapolated using the superposition of these plane waves. First the phases of the plane wave components are shifted according to the position of the extrapolation point. Afterwards the field can be computed at the desired position by summing all plane wave components:

$$P(R, \varphi, \vartheta, k) = \int_0^{2\pi} \int_0^{\pi} w_s(\phi, \theta, k) e^{ikR(\cos \vartheta \cos \theta + \cos(\varphi - \phi) \sin \vartheta \sin \theta)} d\phi d\theta. \quad (5.1)$$

To model a specific combination of pressure and particle velocity in a desired point as captured by a first-order sensor characteristic denoted by $S_\Omega(R, \varphi, \vartheta, k)$,

a factor $g(\phi, \theta, \Phi, \Theta, \beta)$ is included:

$$S_{\Omega}(R, \varphi, \vartheta, k) = \int_0^{2\pi} \int_0^{\pi} w_s(\phi, \theta, k) g(\varphi, \vartheta, \Phi, \Theta, \beta) \cdot e^{ikR(\cos \vartheta \cos \theta + \cos(\varphi - \phi) \sin \vartheta \sin \theta)} d\phi d\theta. \quad (5.2)$$

This factor $g(\varphi, \vartheta, \Phi, \Theta, \beta)$ depends on the angle between the direction of arrival of the plane wave component and the desired steering direction $\Omega = (\Phi, \Theta)$ of the first-order sensor:

$$g(\varphi, \vartheta, \Phi, \Theta, \beta) = \beta + (1 - \beta) [\cos \vartheta \cos \Theta + \cos(\varphi - \Phi) \sin \vartheta \sin \Theta]. \quad (5.3)$$

The value of β allows a transition between omnidirectional and figure-of-eight directivity. Arbitrary sensor characteristics (not limited to first-order characteristics) can easily be obtained by adjusting the weighting factor $g(\cdot)$. In the practical implementation, the double integral in Eq. (5.1) has to be replaced by a finite sum, and corresponding quadrature weights must be included depending on the points for which the directivity function $w_s(\phi, \theta, k)$ is known.

5.3 Spherical Harmonics Extrapolation

Since the spherical harmonic coefficients $A_{nm}(k)$ do not depend on a specific position, they can be used in Eq. (3.1) with $b_n(kR) = j_n(kR)$ from Eq. (3.19), to compute the sound pressure $P(R, \varphi, \vartheta, k)$ in an arbitrary point (R, φ, ϑ) . Hence:

$$P(R, \varphi, \vartheta, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kR) Y_n^m(\varphi, \vartheta). \quad (5.4)$$

The particle velocity vector $\mathbf{V}(R, \varphi, \vartheta, k)$ can be derived as well:

$$ikc\rho_0 \mathbf{V}(R, \varphi, \vartheta, k) = \nabla P(R, \varphi, \vartheta, k). \quad (5.5)$$

The pressure gradient $\nabla P(\cdot)$ is defined in spherical coordinates as (Williams, 1999)

$$\begin{aligned} \nabla P(R, \varphi, \vartheta, k) &\equiv \frac{\partial P(R, \varphi, \vartheta, k)}{\partial R} \mathbf{e}_R \\ &+ \frac{1}{R \sin \vartheta} \cdot \frac{\partial P(R, \varphi, \vartheta, k)}{\partial \varphi} \mathbf{e}_\varphi \\ &+ \frac{1}{R} \cdot \frac{\partial P(R, \varphi, \vartheta, k)}{\partial \vartheta} \mathbf{e}_\vartheta, \end{aligned} \quad (5.6)$$

where $\mathbf{e}_{(\cdot)}$ denote the basis unit vectors. The radial derivative of the sound pressure $P(\cdot)$ in Eq. (5.4) reads

$$\frac{\partial P(R, \varphi, \vartheta, k)}{\partial R} = k \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j'_n(kR) Y_n^m(\varphi, \vartheta). \quad (5.7)$$

The derivative of Eq. (5.4) with respect to the azimuth φ can be written as

$$\frac{\partial P(R, \varphi, \vartheta, k)}{\partial \varphi} = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kR) \frac{\partial}{\partial \varphi} Y_n^m(\varphi, \vartheta), \quad (5.8)$$

where the derivative of the spherical harmonic function is given by Varshalovich et al. (1988)

$$\frac{\partial}{\partial \varphi} Y_n^m(\varphi, \vartheta) = i m Y_n^m(\varphi, \vartheta). \quad (5.9)$$

Similarly, the derivative of Eq. (5.4) with respect to the co-elevation ϑ can be obtained as

$$\frac{\partial P(R, \varphi, \vartheta, k)}{\partial \vartheta} = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kR) \frac{\partial}{\partial \vartheta} Y_n^m(\varphi, \vartheta), \quad (5.10)$$

where the derivative of the spherical harmonics can be calculated with Varshalovich et al. (1988)

$$\begin{aligned} \frac{\partial}{\partial \vartheta} Y_n^m(\varphi, \vartheta) &= \frac{1}{2} \left[\sqrt{n(n+1) - m(m+1)} \right] \cdot Y_n^{m+1}(\varphi, \vartheta) \cdot e^{-i\varphi} \\ &- \frac{1}{2} \left[\sqrt{n(n+1) - m(m-1)} \right] \cdot Y_n^{m-1}(\varphi, \vartheta) \cdot e^{i\varphi}. \end{aligned} \quad (5.11)$$

Finally, the particle velocity $V_\Omega(R, \varphi, \vartheta, k)$ is found by computing the directional derivative of the particle velocity vector $\mathbf{V}(\cdot)$ from Eq. (5.5) with respect

to the direction $\Omega = (\Phi, \Theta)$, (i. e., by computing the scalar product of the pressure gradient and the desired direction vector),

$$V_{\Omega}(R, \varphi, \vartheta, k) = \frac{1}{i k c \rho_0} \nabla P(R, \varphi, \vartheta, k) \cdot \begin{pmatrix} \sin \Theta \cos \Phi \\ \sin \Theta \sin \Phi \\ \cos \Theta \end{pmatrix}. \quad (5.12)$$

Superposition of the sound pressure $P(\cdot)$ in Eq. (5.4) and the particle velocity $V_{\Omega}(\cdot)$ in Eq. (5.12) yields the response $S_{\Omega}(R, \vartheta, \varphi, k)$ of a first-order virtual sensor with steering direction Ω

$$S_{\Omega}(R, \varphi, \vartheta, k) = \beta \cdot P(R, \varphi, \vartheta, k) + (1 - \beta) \rho_0 c \cdot V_{\Omega}(R, \varphi, \vartheta, k), \quad (5.13)$$

where β is the first-order microphone parameter, which is $\beta = 0.5$ in the case of cardioid sensors. Using Eq. (5.13), arbitrary first-order virtual sensors can be calculated within the AAE of the measurement.

5.4 Identity of Plane Wave and Spherical Harmonic Extrapolation

The coefficients of the plane wave decomposition and the spherical harmonic decomposition are mutually dependent. As a consequence it is expected that two methods of extrapolation deliver identical results. This can be verified using the following derivation. The pressure of a sound field in a point (R, φ, ϑ) given by $P(R, \varphi, \vartheta, k)$ can be calculated using Eq. (5.4), which is repeated here:

$$P(R, \varphi, \vartheta, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_{nm}(k) j_n(kR) Y_n^m(\varphi, \vartheta). \quad (5.14)$$

The field can be represented as a number of plane wave components given by $w_s(\phi, \theta, k)$ (c.f. Section 5.2). The directional weight for a single plane wave can be calculated using Eq. 3.12. Inserting this in Eq. (5.14) and integrating over all directions yields

$$P(R, \varphi, \vartheta, k) = \int_0^{2\pi} \int_0^{\pi} w_s(\phi, \theta, k) \left[\sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi i^n \overline{Y_n^m(\phi, \theta)} j_n(kR) Y_n^m(\varphi, \vartheta) \right] d\phi d\theta. \quad (5.15)$$

The term in brackets can be interpreted as a plane wave using Eq. (3.11) and Eq. (3.8) resulting in

$$P(R, \varphi, \vartheta, k) = \int_0^{2\pi} \int_0^{\pi} w_s(\phi, \theta, k) e^{ikR(\cos \vartheta \cos \theta + \cos(\varphi - \phi) \sin \vartheta \sin \theta)} d\phi d\theta, \quad (5.16)$$

which is equal to the plane wave extrapolation discussed in Section 5.2. A simulation with a single plane wave was calculated to illustrate this derivation. Figure 5.2 presents the results of extrapolation based on both principles. A time sequence of band limited impulses with DOA $(\phi, \theta) = (0^\circ, 90^\circ)$ was generated. The spherical harmonic and plane wave decomposition were computed for the sensor array specified in Table 3.4 with $N = \lceil kr \rceil$ and $N \leq 38$ as explained in Section 3.2 and Section 3.4. Only the smaller array radius r_1 was considered and measurement errors were neglected. The spherical harmonic and plane wave extrapolations in Eq. (5.13) and Eq. (5.1) with $\beta = 0$ were used to model figure-of-eight sensors with steering direction towards the DOA of the sound in the first case ($\Omega_1 = (0^\circ, 90^\circ)$) and orthogonal to the DOA in the second case ($\Omega_2 = (90^\circ, 90^\circ)$).¹ The time domain results of the spherical harmonic extrapolation are depicted in the upper half of Figure 5.2(a) and (b) for the steering directions Ω_1 and Ω_2 , respectively. The impulse sequence was band-limited with cutoff frequency $f_c = 7$ kHz, which is below the aliasing frequency f_1 of the sensor array. The size and position of the array are marked by the dashed circle. In both plots, the wave field extrapolation (WFE) is correct inside the sensor array, whereas the extrapolation error outside the array depends on the DOA of the sound and the extrapolation direction. The corresponding results of the plane wave extrapolation illustrated in the lower half of Figure 5.2 show that both extrapolation methods lead to comparable results inside the array. Deviations are due to numerical reasons.

¹For the plane wave extrapolation in Eq. (5.1), the directivity function $w_s(\phi, \theta, k)$ was determined for the directions corresponding to positions of a Lebedev grid with $Q_{\text{Leb}} = 1202$ points. The value of Q_{Leb} is a compromise between computational effort and extrapolation accuracy.

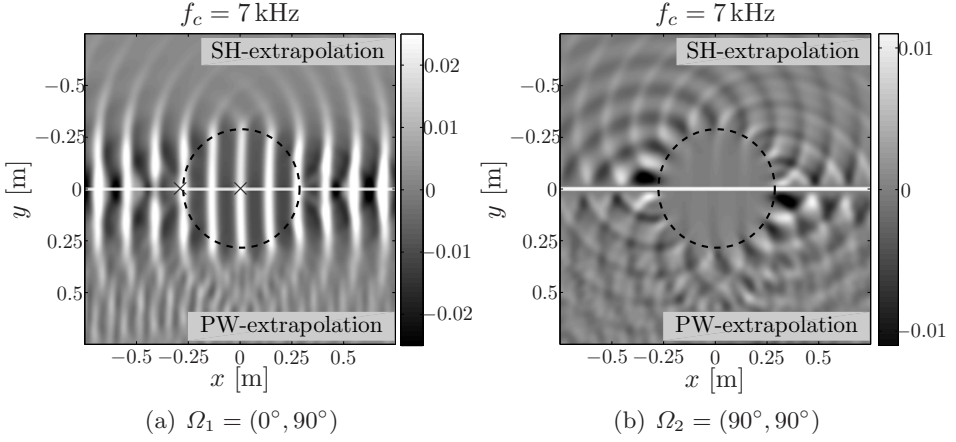


Figure 5.2: The upper halves of the plots demonstrate a SH-extrapolation and the lower half's a PW-extrapolation (time domain). The original sound field was a time sequence of impulses with cutoff frequency f_c . It was captured with the spherical array marked by the dashed circle. The sound field was extrapolated by modeling figure-of-eight microphones with steering direction towards the DOA of the sound in (a) and orthogonal to this DOA in (b).

5.5 Size of the Extrapolation Area and Extrapolation Errors

In Section 3.2 Eq. (3.11) was derived to calculate the field of a plane wave using spherical harmonics. The first sum in Eq. (3.11) is limited to a maximum level N as follows:

$$e^{-i\mathbf{k}\cdot\mathbf{x}} \approx P(\mathbf{x}, k) = 4\pi \sum_{n=0}^N i^n j_n(kr) \sum_{m=-n}^n Y_n^m(\varphi, \vartheta) \overline{Y_n^m(\varphi_0, \vartheta_0)}, \quad (5.17)$$

where (φ_0, ϑ_0) represents the direction of arrival (DOA) of the plane wave and (φ, ϑ) represents the azimuth and co-elevation of the desired position \mathbf{x} . This equation will be studied under various conditions. A plane wave with a temporal frequency of 1 kHz traveling along the x-axis (i.e., in the direction $(0, 90^\circ)$) is considered. The results for different maximum levels N are depicted in Figure 5.3. The axes are normalized by the wavelength λ to make the figures independent of frequency. An adequate approximation of the plane wave can

be observed within a limited area around the origin. The size of the area is increasing with the the level N . In Figure 5.3(a) to 5.3(c) the pressure fields in the (x, y) -plane are depicted, the Figures 5.3(d) to 5.3(f) present the (y, z) -plane. Since the plane wave is traveling along the x -axis, the pressure is constant in the (y, z) -plane. The results again show that pressure can only be approximated exactly for a limited region around the origin. An empirical rule to estimate the area of accurate extrapolation (AAE) can be given by

$$d \approx \frac{N\lambda}{2\pi}, \quad (5.18)$$

where d represents the distance from the origin where the pressure has decreased by 3 dB (Rafaely, 2004). It can be concluded that the coefficients of a higher level n contain the information of the sound field within a longer distance from the origin. The following developments are based on this insight. In view of the previous sections it is important that the AAE is given by the radius of the array when $N_{\max} = \lceil kr \rceil$. The AAE is decreased, if N_{\max} is limited further. The influence of measurement errors and further extrapola-

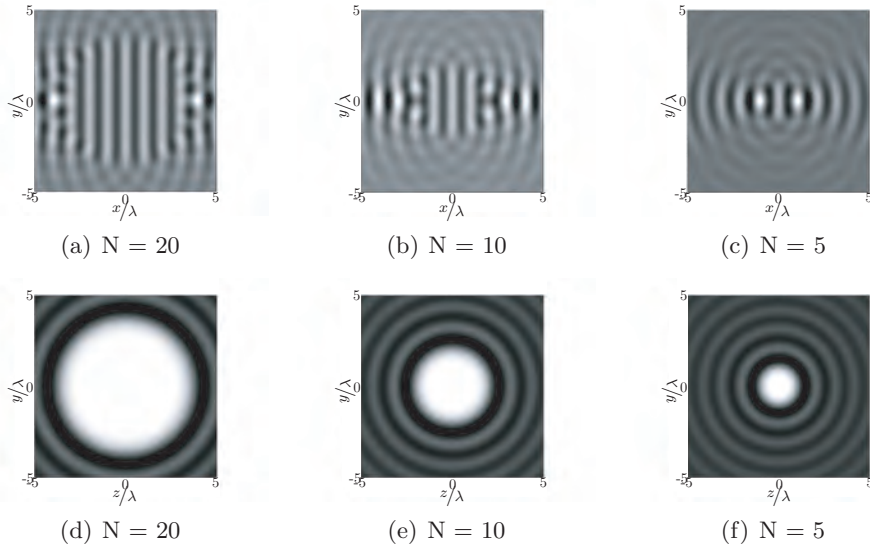


Figure 5.3: Level-limited pressure field of a 1 kHz plane wave traveling along the x -axis in linear scale (a,b,c show the (x,y) -plane; d,e,f show the (y,z) -plane).

tion problems are examined now. Since the spherical Bessel function $j_n(kR)$ in Eq. (5.14) becomes approximately zero for orders $n > kR$ very quickly, only

coefficients $A_{nm}(k)$ up to the maximum order $N \approx kR$ are required to extrapolate a sound field with little error to a distance R from the origin. Therefore, the largest distance R of an accurate WFE can be approximated with

$$R \approx \frac{N}{k}. \quad (5.19)$$

As presented in the previous section, the AAE equals the size of the sensor array when setting the maximum order to $N \approx kr$, where r is the array radius. If N is constant, the AAE will decrease linearly with increasing wavenumber k . To study the effect of spatial aliasing on the WFE, Figure 5.4(a) shows the

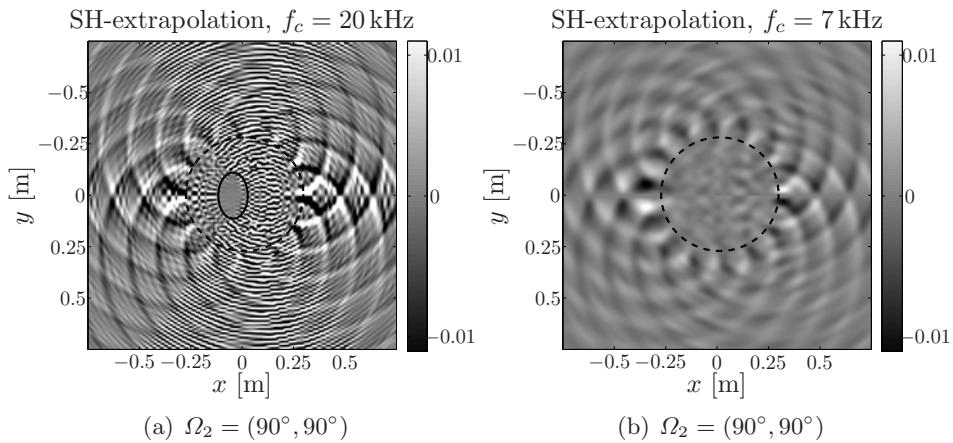


Figure 5.4: SH-extrapolation of a time sequence of impulses with cutoff frequency f_c . The extrapolation was computed by modeling figure-of-eight microphones with steering direction orthogonal to the DOA of the sound. Plot (a) shows the influence of spatial aliasing and plot (b) shows the influence of microphone noise and of position inaccuracies of the array.

extrapolation based on spherical harmonics from the simulation of the previous section, this time for a cutoff frequency $f_c = 20$ kHz, which is above the aliasing frequency f_1 . The extrapolated figure-of-eight microphones with steering direction Ω_2 are directed orthogonally to the DOA of the sound. Comparing the plot with the aliasing-free situation in Figure 5.2(b), strong aliasing effects can be observed, also inside the sensor array. Nevertheless, the WFE can be computed correctly within a specific area near the origin (see the inner circle). The results reveal that spatial aliasing mainly affects the spherical harmonic coefficients $A_{nm}(k)$ of higher orders n , which contain information on the sound field for larger distances from the origin (Daniel et al., 2003). Figure 5.4(b)

illustrates the results for $f_c = 7$ kHz when including the measurement errors listed on page 82. No AAE can be observed, indicating that the coefficients $A_{nm}(k)$ of lower orders n are also corrupted by the measurement errors.

5.6 Coherence in Extrapolated Fields

Since the sound field of a room consists of discrete reflections in the early time after the direct sound, it becomes more and more diffuse due to high order reflections in the later part. To study the influence of extrapolation on the late part of an acoustic field, it is important to study the coherence in extrapolated fields. This parameter characterizes the diffuse or late part. The coherence of two points in a diffuse sound field depends on the parameters for the measurement. As shown in the previous section, the order N is one of the most important parameter. Avni and Rafaely (2009) have presented a derivation for a monochromatic diffuse field resulting in a spatio-temporal correlation coefficient given by

$$\rho(k, r) = \frac{\sum_{n=0}^{\infty} (-1)^n (2n+1) |4\pi i^n j_n(kr)|^2 \cos(\omega\tau)}{\sum_{n=0}^{\infty} (2n+1) |4\pi i^n j_n(kr)|^2} \quad (5.20)$$

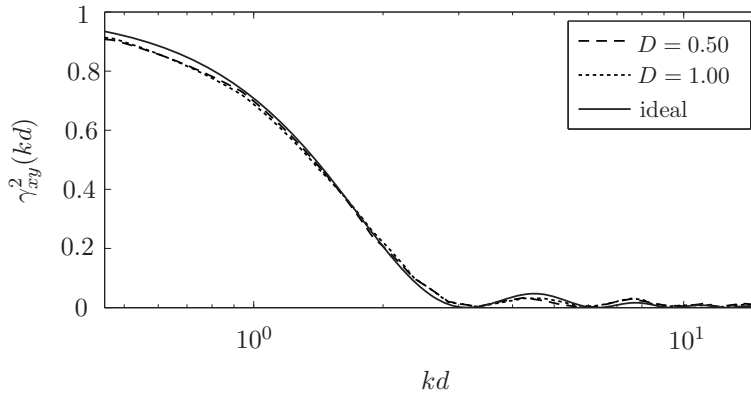
for the special case of two measurement points at the coordinates $(r, \frac{\pi}{2}, 0)$ and $(r, \frac{\pi}{2}, \pi)$. As a result the upper frequency for which the coherence of a field can be reconstructed again can be approximated by $N = kr$ using $\omega = \frac{Nc}{r}$. To study the influence of aliasing and extrapolation on the coherence of the extrapolated field, a broadband diffuse sound field is approximated in the frequency domain. A simulation was made by superimposing plane waves arriving from $Q = 1202$ uniformly distributed directions, where the DOAs are found from a Lebedev grid. Each plane wave carries a zero-mean white Gaussian noise signal. The field is used to simulate the responses of different spherical microphone arrays. The wave field is extrapolated to different positions based on these simulations. The influence of the extrapolation distance and of spatial aliasing on the coherence between two extrapolated points is studied by comparing the simulation with the ideal coherence function as given in Section 2.7.

5.6.1 Influence of the Extrapolation Distance

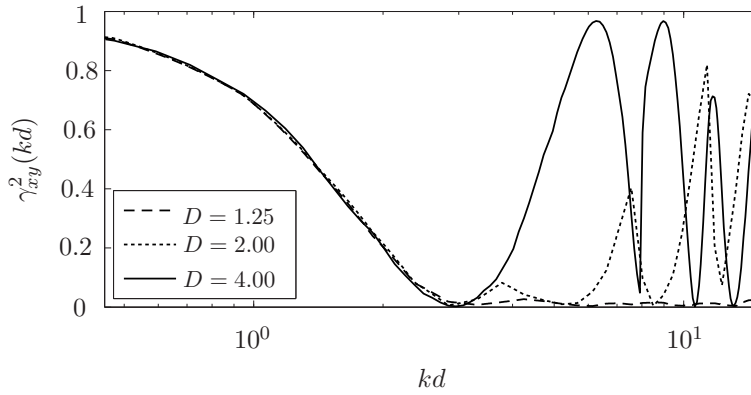
To investigate the effect of the extrapolation distance on the spatial coherence of the extrapolated sound, any influence of spatial aliasing must be eliminated. For this reason, the coefficients $A_{nm}(k)$ are computed for each plane wave and frequency using Eq. (3.21) with $N = \lceil kr \rceil$. Each set of coefficients is then multiplied by a normally distributed random magnitude and a uniformly distributed random phase. After adding up the results for all waves, the WFE is computed using Eq. (5.13). For the analysis, the pressure in two points in opposite directions at a distance r_0 from the origin was calculated. Hence, the distance between both sensors is $d = 2r_0$. Due to the choice of the maximum order N , the wave field extrapolation (WFE) can be computed correctly for all distances $r_0 \leq r$. The spatial coherence $\gamma_{xy}^2(k, d)$ between the outputs of both sensors is determined using Eq. (2.73). Figure 5.5 shows the results for different ratios $D = r_0/r$, as a function of kd , compared to the ideal result computed with Eq. (2.76a). The plots are an average over 10 realizations of the diffuse sound field. For $D \leq 1.00$ (i. e., when extrapolating to positions inside the array) the coherences obtained from the extrapolated signals agree with the values of an ideal diffuse field. A similar result is also attained for $D = 1.25$. Thus, the WFE does not impair the diffuse characteristic of the sound field. When extrapolating to positions far outside the array ($D \geq 2.00$), strong coherences between both virtual sensors do arise at higher values of kd . In this range of kd , the sound field is no longer incoherent as desired.

5.6.2 Influence of Spatial Aliasing

Next, the influence of spatial aliasing is studied using the geometry given in Figure 5.6. To study the influence of spatial aliasing on the spatial coherence function in an extrapolated field, a microphone array with radius r and with sensors located on a Lebedev grid is simulated. The plane waves are generated in the frequency domain by multiplying each frequency bin of a white unit spectrum with a normally distributed random magnitude and a uniformly distributed random phase. The waves are then shifted to each sensor position by adjusting the phase and are weighted with the cardioid microphone characteristic. After superimposing all waves, the spherical harmonic decomposition (SHD) and WFE is computed for omni-directional virtual sensors using Eq. (5.14) and Eq. (5.13), respectively. In the first example two extrapolation positions P_1 and P_2 are considered (see Figure 5.6(a)) located at opposite di-

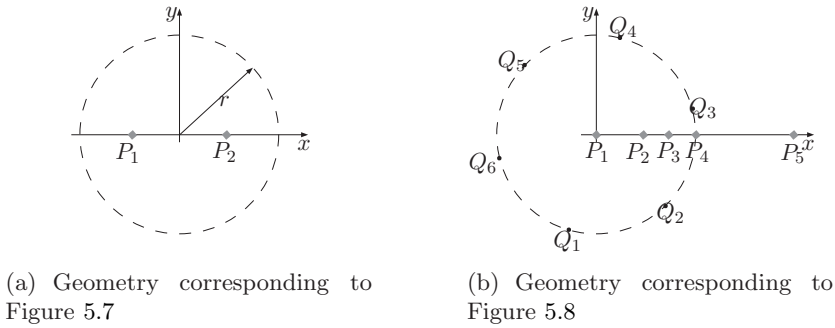


(a) $D \leq 1$



(b) $D > 1$

Figure 5.5: Spatial coherence for two omni-directional virtual sensors. Spatial aliasing is not present. The WFE was computed for different ratios $D = r_0/r$, where r_0 is the distance of the extrapolation points from the origin and r is the array radius. The distance between the virtual sensors is $d = 2r_0$.

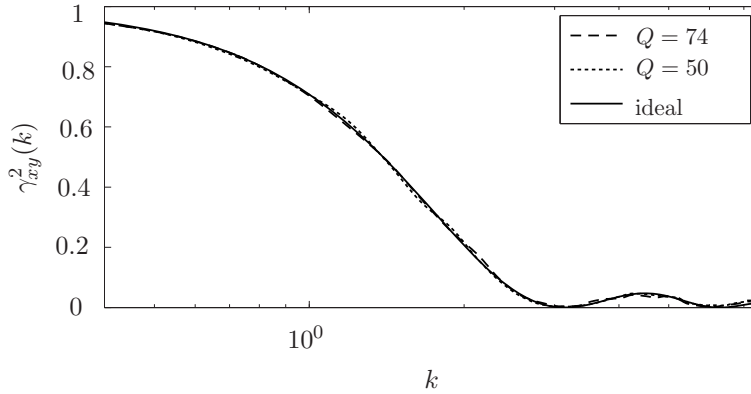


(a) Geometry corresponding to Figure 5.7

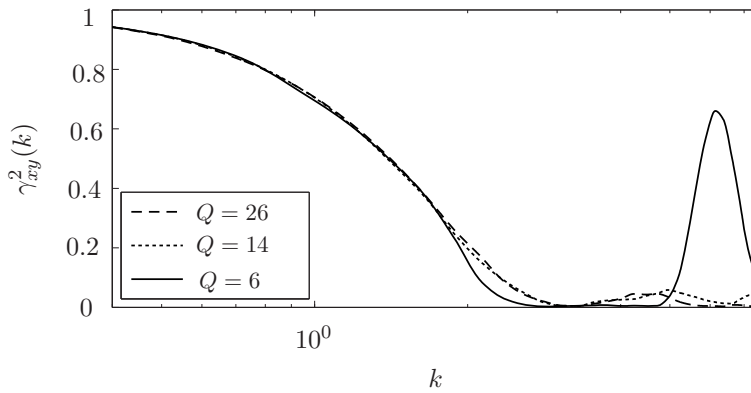
(b) Geometry corresponding to Figure 5.8

Figure 5.6: Geometry used for the simulation of coherence. The extrapolation positions are indicated by P_1 to P_5 . The array radius in (a) is given by r while the number of sensors is varied. In (b) a limited number of sensors was used, indicated by Q_1 to Q_6 and the field is extrapolated to a coincident microphone position at P_1 to P_4 .

rections at a distance r_0 from the origin. The ratio between the extrapolation distance and array radius is $D = r_0/r = 0.75$. Figure 5.7 presents the coherence between P_1 and P_2 , plotted as a function of k for arrays with a different number Q of sensor elements. The results for an ideal diffuse field, computed with Eq. (2.76a), are depicted for comparison. When using a sufficiently high number of sensor elements ($Q = 74$, array radius $r = 15$ cm), a strong similarity to the ideal result can be observed. For lower numbers of sensor elements, significant inaccuracies do arise. The coherence function is then characterized by high values for specific k . Figure 5.8 shows the coherence between two perpendicular velocity sensors in a coincident layout. The WFE was computed for different ratios $D = r_0/r$, where r is the array radius and r_0 is the distance of the virtual sensor-configuration from the origin. In Figure 5.8(a) a microphone array with six sensors was simulated, including spatial aliasing. Due to aliasing, the coherence function for positions with longer distance to the origin is degraded. In the case of an aliasing-free simulation (see Figure 5.8(b)) the degradation of the coherence function will not occur until the position is outside the microphone array radius ($D = 2$). The reason is that coefficients are corrupted by aliasing, which are required to allow a sufficient large extrapolation distance. Note that in the current plots a relatively small frequency range is shown.

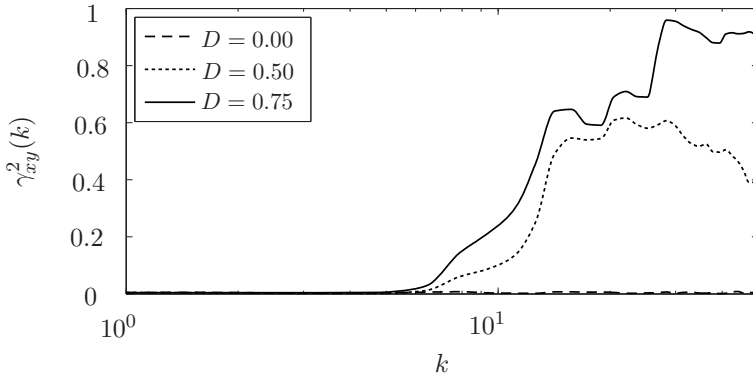


(a)

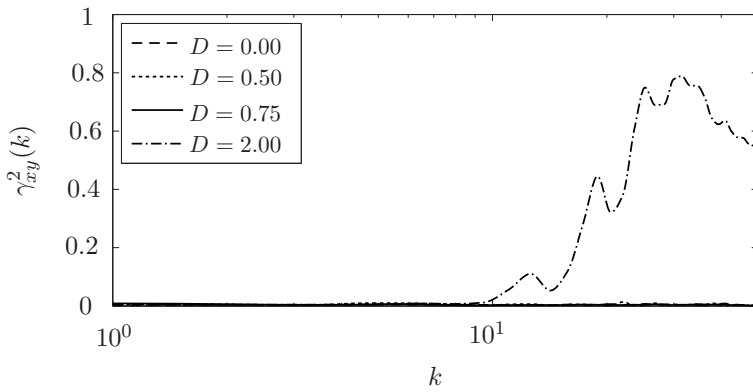


(b)

Figure 5.7: Spatial coherence between two omnidirectional virtual microphones with distance d between each other. Spatial aliasing is included. The WFE was computed for arrays with different number of sensor elements Q . The ratio between the distance r_0 ($= d/2$) of the virtual sensors from the origin and the array radius r is $D = r_0/r = 0.50$.



(a) Including spatial aliasing.



(b) No spatial aliasing.

Figure 5.8: Spatial coherence between two virtual, perpendicular, coincident velocity sensors. The WFE was computed for different ratios $D = r_0/r$. r_0 is the distance of the virtual sensor configuration from the origin and r is the array radius. A array with $Q = 6$ microphones was simulated.

5.7 Summary

In this chapter the extrapolation of measured wave fields was studied. An important result is that the information for larger distances from the origin is stored in the spherical harmonic coefficients of higher level. The area of accurate extrapolation (AAE) was defined and an equation to derive it for error-free coefficients was given. Methods for extrapolation based on spherical harmonics as well as on plane wave decomposition were studied and compared. It was shown that the methods are equivalent and are subject to the same restrictions due to limitation of the maximum level of the spherical harmonics employed and size of the measurement aperture. The influence of spatial aliasing and measurement errors on the extrapolation was demonstrated based on spherical harmonic extrapolation. Furthermore, the influence of errors on extrapolated diffuse fields have been studied in terms of coherence. It was shown that extrapolation and aliasing errors degrade the coherence function of extrapolated fields. The aliasing error can be recognized in case of an insufficient number of grid points as a degradation of the coherence function. The extrapolation error is present when the extrapolation distance is long. The findings of this section are relevant for the adaptation of spherical array measurements to different reproduction systems, as will be discussed in Chapter 7.

Chapter 6

Interaction

6.1 Introduction

In this chapter it is investigated how measurement data can be presented visually. New principles by which the sound designer can interact with such data are developed. This is needed in every kind of sound production process because the aim is not to auralize a given measured acoustic field, but to produce an optimal acoustic impression for a piece of music or a motion picture sound track. In Chapters 3 and 4, techniques for plane wave decomposition (PWD) of room impulse responses based on spherical array measurements were given. All visualization and interaction principles described in this section will be applied to plane wave decompositions. The visualization is presented such that the spatial impression is emphasized. The visualization is the basis for applying direct interaction techniques which allow direct modification, as defined by (Frohlich, 1993), of such visualization by the user. An important goal is to develop visualization and interaction methods which are independent of the reproduction system. This chapter corresponds to block (5), (6), (7), and (8) in Figure 6.1. To make the interaction fast and most effective for the user, psychoacoustic knowledge will be applied to optimize the data representation and interaction. In contrast to other works, the aim is not to compress the amount of data as proposed for impulse responses in (Hulsebos, 2004) or implemented in several systems for audio coding (Kahrs and Brandenburg, 1998) (Noll, 1997). The main goal is to develop representations which have a strong correlation with perception. This can be achieved by removing perceptually irrelevant details and optimize parameter ranges (e.g., the range of time- and frequency resolution or the amplitude range present in a visualization). In other words, on the visualization side, psychoacoustic knowledge is used to optimize the graphical representations. On the interaction side, it is used to

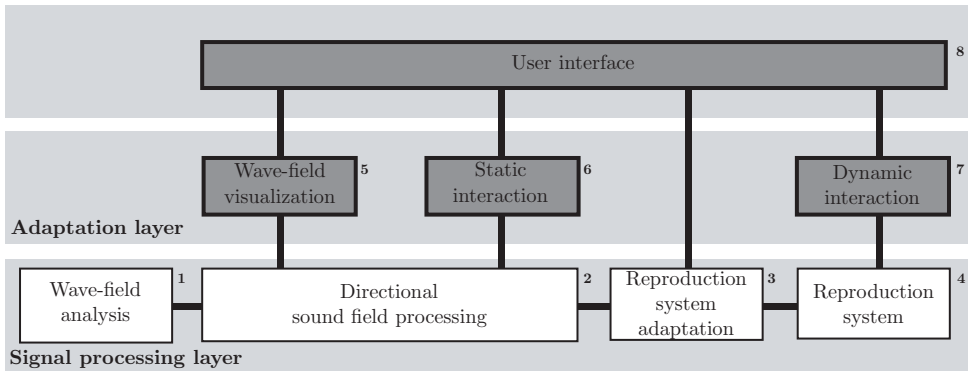


Figure 6.1: Basic block diagram of the system developed in this thesis. The parts discussed in this chapter are emphasized.

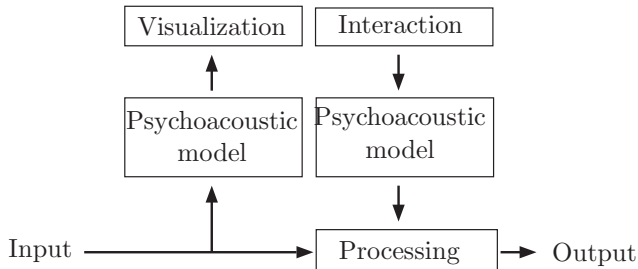


Figure 6.2: Integration of psychoacoustic knowledge into the sound design process.

make the results of any modification directly and clearly audible. The concept is depicted in Figure 6.2.

6.1.1 Basic Chain of Interaction

Figure 6.3 represents the data flow during the interaction and adaptation process. The *high resolution data* are adapted to the desired reproduction system. The result is stored as *reproduction data*. This step is studied in Chapter 7. The interaction process is divided into two parts, which termed static interaction (SI) and dynamic interaction (DI).

Static interaction is based on analyzed *high resolution data* from an array measurement. The generation of such data based on spherical array measure-

ments was described in Chapter 4. This type of data is the preferred storage format during the interaction process for modified rooms (denoted as Room presets). The static interaction (SI) process is independent of the reproduction system in terms of usability and signal processing.

Dynamic interaction During the reproduction process the user can manipulate the characteristic of the sources. This process is named dynamic interaction (DI) because it depends on the source parameters and has to be changed in real time during the reproduction. The DI process is independent of the reproduction system in terms of usability but not in terms of signal processing.

6.1.2 Outline of the Chapter

The first step in the development of interaction principles is to study visualization techniques for room impulse responses. A taxonomy of visualization is given in Section 6.2. The SI process as defined in the previous section is investigated in Section 6.3. Both sections start with single impulse responses and extend the techniques to direction-dependent multi-trace impulse responses. The DI process is studied in Section 6.4. In the last Section 6.5, two user interfaces developed as part of this work are presented. These are used to implement the interaction methods based on a standard monitor, keyboard and mouse environment as well as on novel interface techniques.

6.2 Taxonomy of Room Impulse Response Visualization

The visualization of impulse responses delivers the basis for a direct interaction by the user. For this reason a taxonomy of visualization types is given in this section. The taxonomy is based on diagram types. The classification is developed from single impulse responses (Section 6.2.1), to multiple, or direction-dependent visualizations (Section 6.2.3).

Simulation used for visualization For the illustration of different diagram types, a mirror image source model (MISM) was used. A directional source

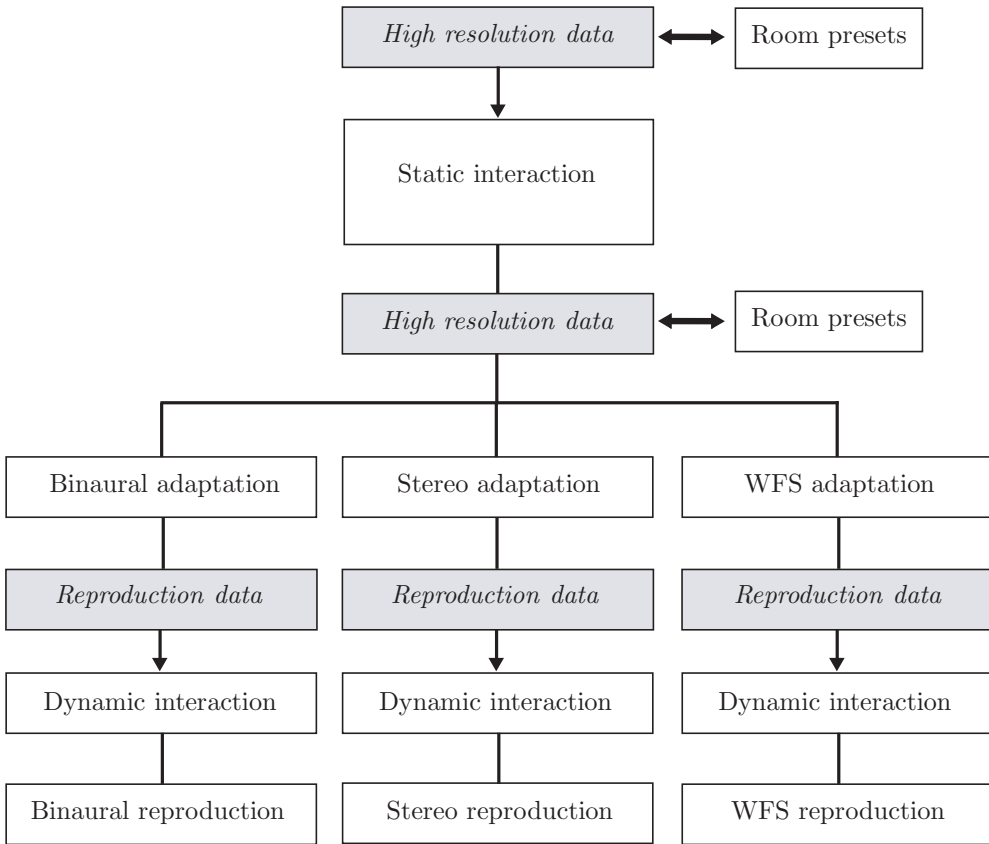


Figure 6.3: Block diagram of the complete interaction process. The different data sets to store are indicated as gray blocks. The white blocks indicate signal processing and audio reproductions steps.

No	Diagram	x axis	y axis	$y(x)$
1	Amplitude	time	amplitude	$s(t)$
2	Energy curve	time	energy	$s(t)^2$
3	Envelope	time	amplitude	Eq.(6.1)
4	Energy decay curve	time	inverse energy integral	Eq.(6.3)
5	Frequency curve	frequency	magnitude	$ S(\omega) $

Table 6.1: Axis types for single impulse response two-axis visualizations.

in a rectangular room is considered. The source has a directivity of a broad cardioid and is oriented to the left wall at an angle of 45° . The walls have different frequency-dependent reflection coefficients to illustrate the potential of direction-dependent room impulse response analysis and visualization. The left hand side wall has a high-pass, the right hand side wall has a low-pass characteristic. The front and back wall as well as the floor and the ceiling have a broadband characteristic but large and small reflection coefficients, respectively. The PWD is calculated in an idealized form:

1. The impulse response for each reflection is calculated, considering delay, distance attenuation, air damping, and the frequency-dependent reflection coefficients.
2. The angle of incidence for each reflection is rounded to a grid of 36 azimuth directions and a single co-elevation angle.
3. All reflections are summed to these 36 direction-dependent impulse responses.

The result of this process is equal to projecting all directions to the horizontal plane comparable to a measurement with a circular array.

6.2.1 Visualizations of Single Impulse Responses

Two-axis Diagrams

Table 6.1 presents an overview of visualizations for single impulse responses using two-axis diagrams.

Waveform representation The simplest way to visualize an impulse response is to plot it directly in the time domain using a two-axis diagram as given in Figure 6.5(a). From a perceptual point of view, the information value in this visualization is not very high. The reason is that the fine structure of the impulse response will not be perceived directly. Derivatives of such representations can be useful if insignificant details are excluded beforehand. One can consider several models to rectify and smooth the impulse response. Two common methods are given in the next two paragraphs.

Envelope representation A well-defined method is the calculation of an envelope based on the Hilbert transform of the impulse response (Kuttruff, 1991). The envelope $e(t)$ of a given impulse response $h(t)$ can be calculated using

$$e(t) = \sqrt{h(t)^2 + \mathcal{H}\{h(t)\}^2}, \quad (6.1)$$

where the Hilbert transform $\mathcal{H}\{\cdot\}$ is defined by

$$\mathcal{H}\{h(t)\} \equiv \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{h(t-t')}{t'} dt'. \quad (6.2)$$

An example can be found in Figure 6.5(b). The envelope is calculated using Eq. 6.1 and the result is presented in a logarithmic scale. Please note that due to the ideal calculation the values are zero between the reflections. The graph is truncated at -60 dB.

Energy decay curve Another two-axis representation of the RIR is the energy decay curve (EDC), also called the Schroeder integral (Schroeder, 1965). The EDC $s(t)$ of an impulse response $h(t)$ can be calculated using the Schroeder integral given by

$$s(t) = \int_t^{\infty} h^2(\tau) d\tau. \quad (6.3)$$

The logarithmic Schroeder integral of the simulated impulse response is depicted in Figure 6.5(c). It is also used to calculate the reverberation time using linear curve fitting.

No.	Diagram	x axis	y axis	z(x,y)
1	Energy decay relief (EDR)	time	frequency	$P_{EDR}(t, \omega)$
2	STFT	time	frequency	$ P(t, \omega) $

Table 6.2: Axis types for single impulse response three-axis visualizations.

Other representations The representation of the RIR in the frequency domain is also possible. The representation of the single spectrum is not very informative from a sound design perspective because the impulse response is always perceived as a time-variant filter and not as an average frequency distribution. For this purpose the two-axis representations have to be augmented with another dimension.

6.2.2 Three-axis Diagrams

The logical augmentation of the diagrams presented in the the last section is the utilization of frequency information in a time dependent representation yielding three axis diagrams. An overview of three axis diagrams is given in Table 6.2.

Time-frequency representations Time-frequency analysis can be performed by applying a short time Fourier transform, as given in section 2.5, to the representations 1 to 4 in Table 6.1. An example of a STFT-representation of a RIR is given in Figure 6.5(d). Now the frequency distribution of the reflections can be studied. Due to the nature of the STFT, either frequency or time resolution is always limited at the expense of the other. One might consider using other analysis tools and filter banks. For example a third-octave filter bank can be applied. This kind of filter bank is very common in the field of audio processing and is not far from the frequency resolution of the human hearing system. Another possibility is the application of filter banks which model the human perception (Moore, 1995) (e.g., given by (Stanley, 1993)). A detailed study of such methods is beyond the scope of this thesis. The use of a simple hearing model for room impulse response visualizations was proposed by Lokki and Karjalainen (2000) and Merimaa et al. (2001). These are based on the use of a gammatone filter bank and an adapted integration window. The auditory model used is shown in Figure 6.4. van Dorp Schuitman and de Vries (2008) used the processing of psychoacoustic models in terms of denoising of

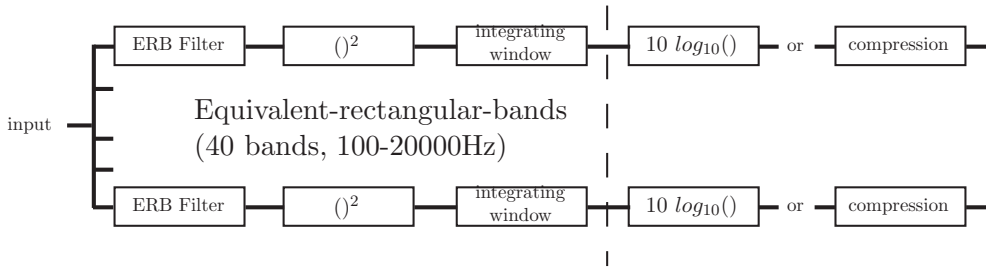


Figure 6.4: Psychoacoustic motivated analysis method for RIRs used in (Lokki and Karjalainen, 2000).

impulse responses. Perceptually irrelevant noise in the impulse response is one possible reason for the spatial fluctuations of the measure of spaciousness parameter reported in (de Vries et al., 2000), in addition to the interference of wave fronts in a single measurement point.

Energy decay relief Another useful three-axis representation is the energy decay relief plot as proposed by Jot (1992); Jot et al. (1997). This can be interpreted as a frequency-dependent version of the energy decay curve. The general time-frequency representation of a impulse response $h(t)$ is given by $P(t, \omega)$. The EDR denoted as $P_{EDR}(t, \omega)$ is defined by Jot et al. (1997), as:

$$P_{EDR}(t, \omega) = \int_t^{\infty} P(\tau, \omega) d\tau \quad (6.4)$$

This representation is useful for the frequency-dependent analysis of the late reverberation. An example can be found in Figure 6.5(e). The different reverberation times for the high frequency components in comparison to the low-frequency components can clearly be seen.

Three-axis diagrams including directional information When a sound field is analyzed in a direction-dependent way as described in Chapter 3, this information can also be applied to the time-frequency representation. Lokki and Karjalainen (2000) proposed an extension using small arrows in each frequency bin to indicate the direction of incidence of the energy component. Figure 6.5(f) presents an example. While this representation gives a lot of information in a single diagram, the identification of the directions of incidence

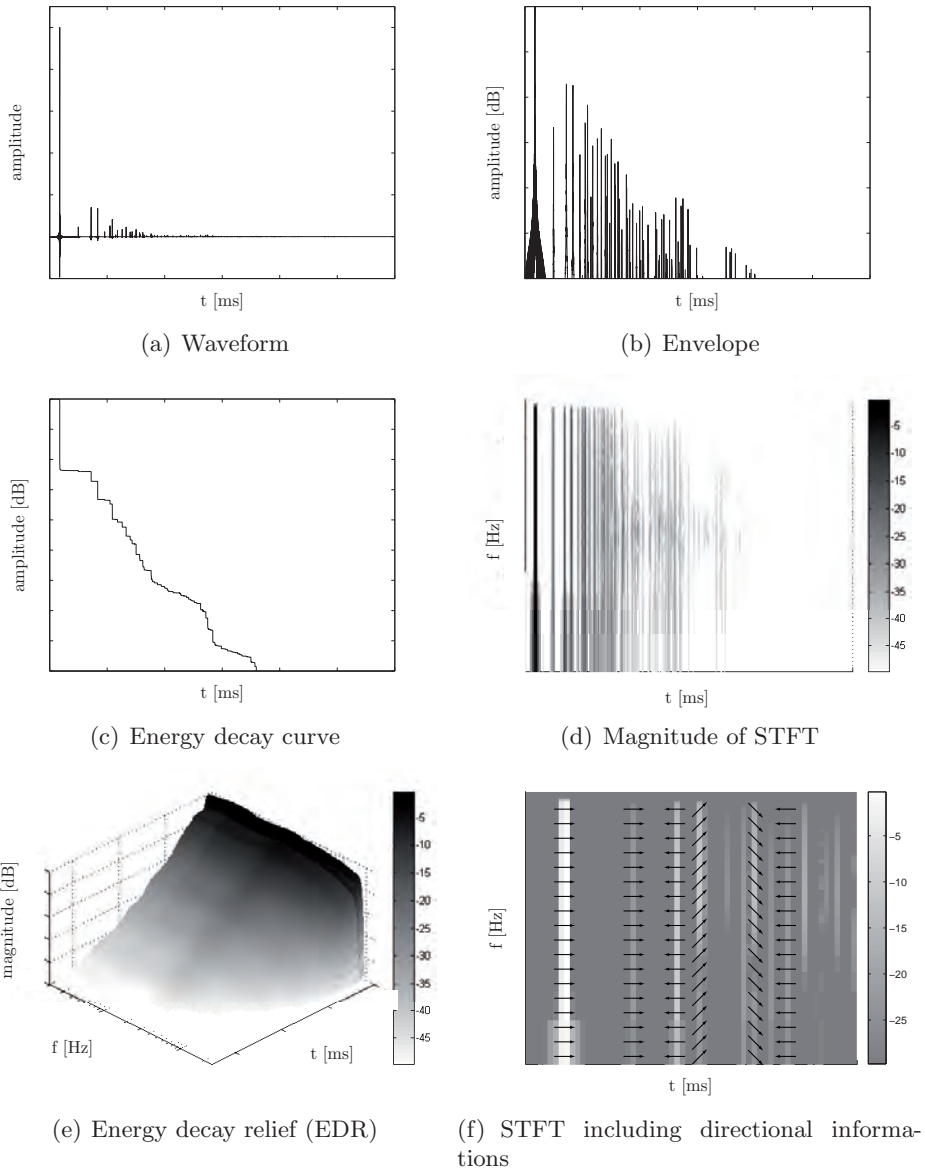


Figure 6.5: Qualitative examples of diagram types for impulse response visualization using axis types given in Tables 6.1 and 6.2. Please note that the axis scales have been adapted to allow an adequate representation of each diagram.

of single events can be difficult. A more direct connection between the angles of incidence and the visualization is preferred. Examples will be developed in the following sections.

6.2.3 Visualization of Multiple Impulse Responses

Plane wave decomposition enables the directional analysis of a sound field. It is desirable that the spatial information is included in the visualization. In this section new methods are given which are developed in view of the direct interaction.

Multi-trace representation A common visual representation method is to plot impulse responses as a waveform representation according to their angle of incidence. The result is named *multi-trace impulse response*. These can give information about the angles of incidence of different reflections. Easier to understand are sound pressure measurements along a linear array, which show the incoming wave front directly. For this reason such representations can be found in the analysis of wave fields (e.g.,(de Vries et al., 1997))¹.

Plane wave decomposition representation In the case of a plane wave decomposition as given for spherical arrays in Chapter 3, the sound field can be analyzed with high spatial resolution. In this case, one axis of a three-axis representation can be used to represent the angle of incidence. The remaining axis represents the time or the frequency. Should only a single frequency be analyzed, two axes can be used to give the azimuth as well as co-elevation angle. These representations have been used extensively in the previous chapters. For an interactive auralization application, the single frequency representation is not very useful. Since an auralization application will always a broadband application. For this reason, only one axis can be used for the direction. The amplitude can be coded as color or height. Examples are given in Figure 6.6. To make such a diagram more intuitive, it is useful to arrange the RIRs corresponding to their direction of incidence.

¹A very similar visualization can be found in the early work of (Sabine, 1964). It was generated by impulse response measurements, but it also visualized the wavefronts of direct sound and corresponding reflections.

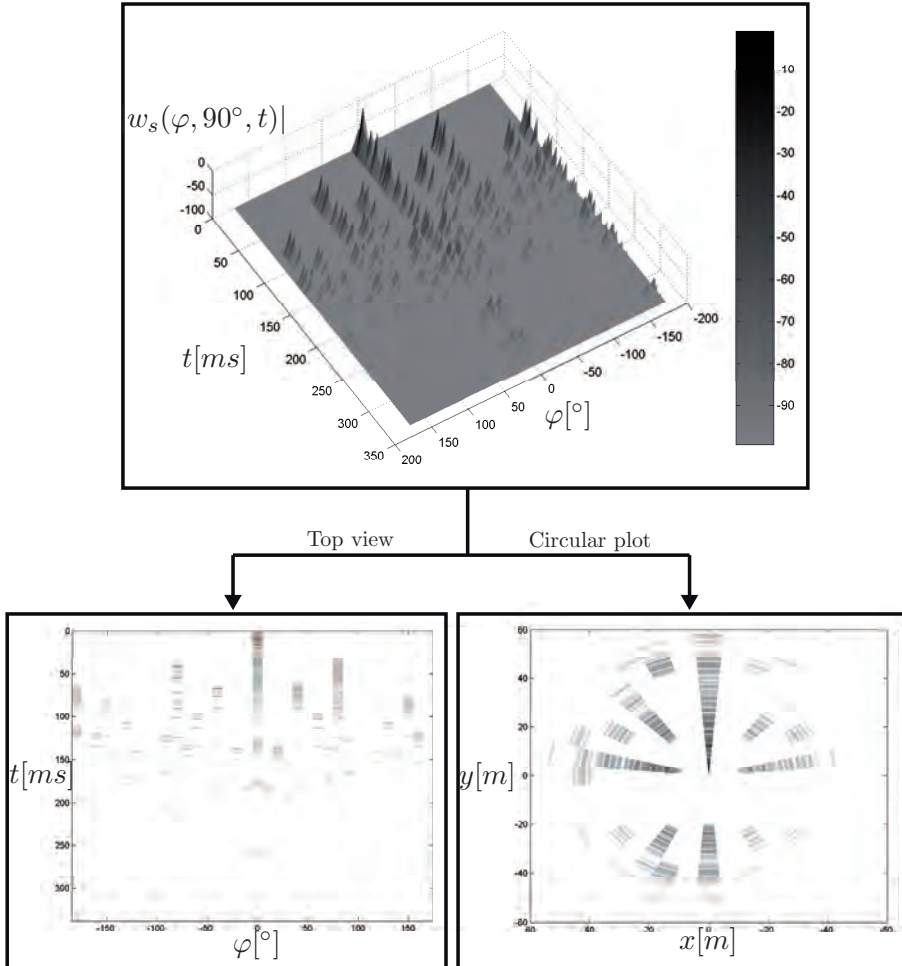


Figure 6.6: Example of plane wave decomposition representations. In the upper and left visualization, a multi-trace representation is used. The direction is represented as azimuth angle in the x-axis. The right visualization uses the azimuth angle as polar coordinates to represent the direction of incidence more intuitively.

Volume-data representation In the case of an interactive visualization² it is advantageous to use volume-data representations. In such a representation each point in a three-dimensional space is characterized by a value. The user can navigate through such data by placing slices in the dataset. The slice represents the values for this section using color coding. Furthermore, iso-surfaces³ can illustrate the areas of the same value in the volume. Figure 6.7 gives an schematic view of such a diagram. The user can modify the slices for the different dimensions using an interaction device as discussed in Section 6.5.2. An example of a volume data representation can be found in Figure 6.8. In this example the impulse response of an acoustically highly asymmetrical room (see Figure 6.8(a)) was measured. This property can clearly be identified in the measurement data visualization as given in Figure 6.8(b). An alternative version of the volume data representation is the use of a spherical volume element. The radius of the sphere represents the time-axis. For each time frame the magnitude values of the PWD are color-coded on a spherical surface. Figure 6.9 presents an example. In this case the anechoic room measurement from Section 3.7 is visualized. A different variant can be generated by selecting a specific angle window and representing the complete direction-dependent impulse response. This can be interpreted as a modern version of the *hedgehog plot* used in the work of (Thiele, 1953), more recently republished in (Rieländer, 1982). The advantage, beside the modern visualization, is that the complete impulse response for a specific direction can be examined and not only the energy of one time frame.

Three-axes Representation

An alternative way of including directional information is the spatial arrangement of three-axis diagrams representing the angle of incidence for the different RIRs. This is useful if a limited number of directions are to be displayed. Figure 6.10 presents such a diagram for the ideal PWD representation given in the previous images. The number of directions is set to eight. In this representation, the spatial details of the simulation are clearly visible. The reflections from different directions as well as their frequency distribution can be identified. The direct sound can be seen in the upper diagram. The different side wall properties such as the high-pass characteristic of the right wall and the low-pass characteristic of the left wall can be noticed. This representation will

²Visualizations in which a user can navigate.

³Surfaces of constant values.

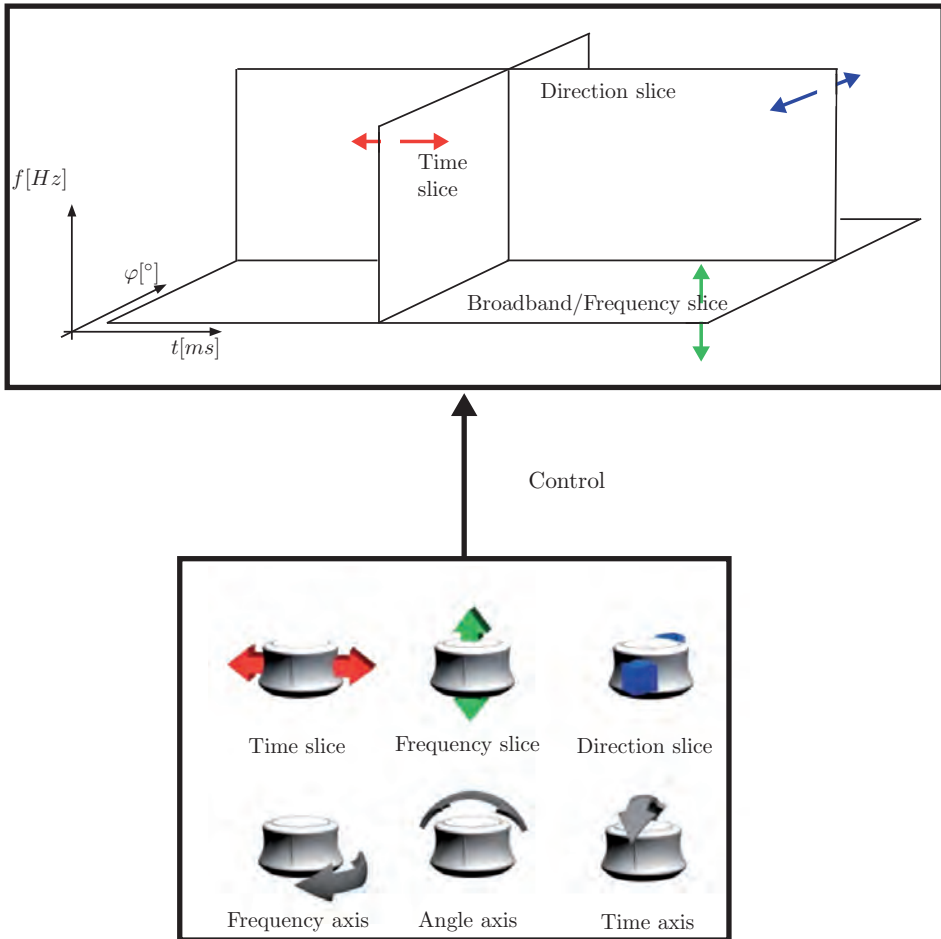
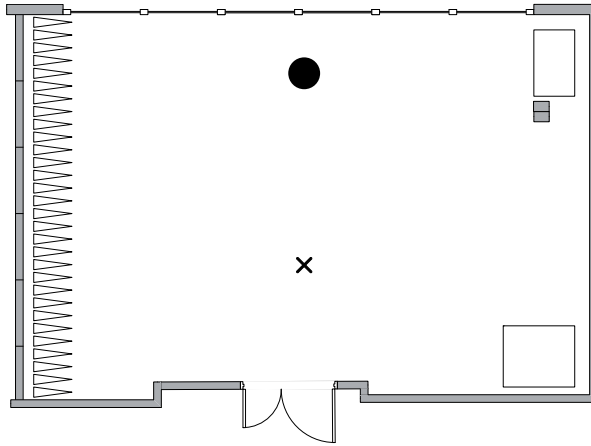
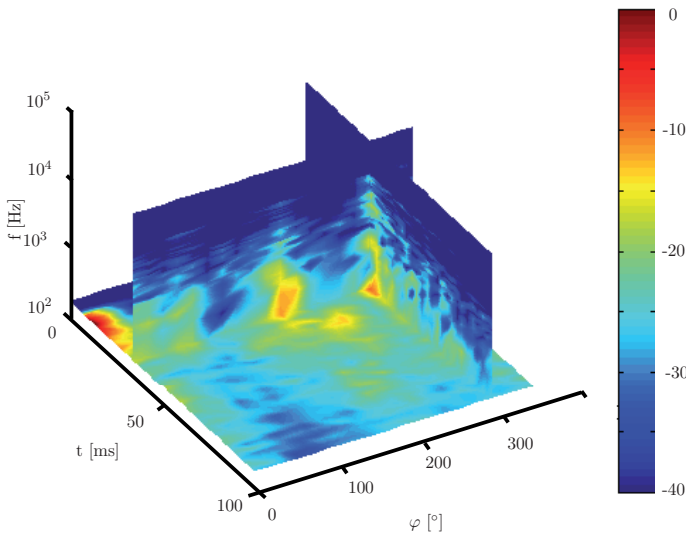


Figure 6.7: Example of a three-axis volume data representation with three slices. The lower surface is used for broadband representation. This enables orientation while moving the time and direction slices. In the lower part a possible controller and its connection to the visualization is shown.

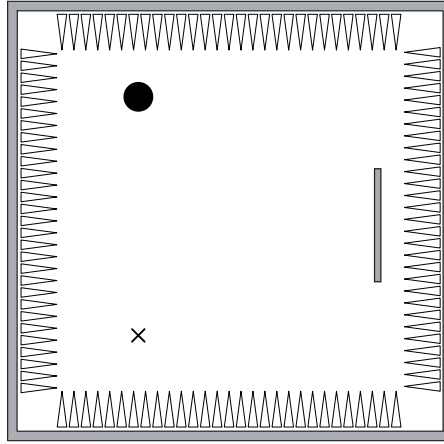


(a) Geometry of the measured room.

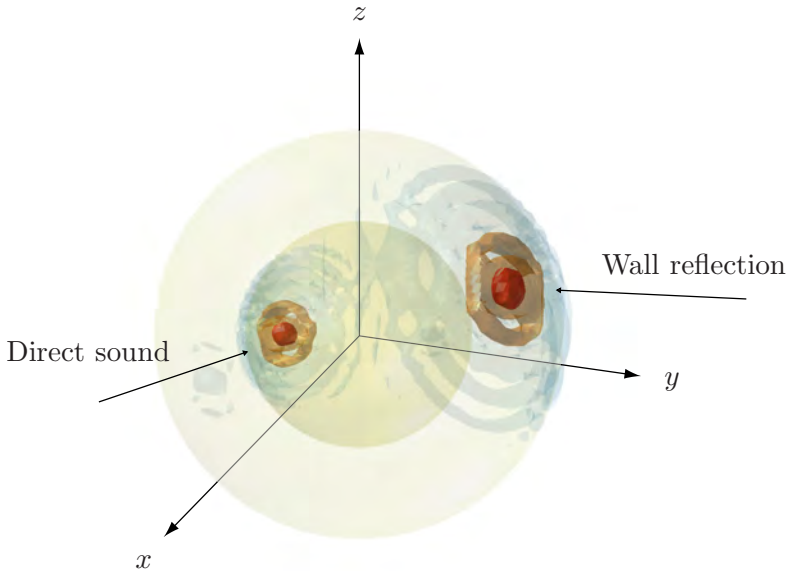


(b) Volume data representation.

Figure 6.8: Example of Volume data representation. The room shown in (a) was measured with a microphone array. The result is shown in (b) as volume data representation. The difference between the left and right wall can clearly be identified. While the left wall (90°) was damped, the right wall (270°) was high reflective. The volume data show a measurement slice of $(\varphi, 90^\circ, t, f)$.



(a) Geometry of the measured room.



(b) Volume data representation in spherical coordinates.

Figure 6.9: Plane wave decomposition of a sound field composed of a direct sound and a wall reflection. The direct sound arrives with a DOA $(\varphi_1, \vartheta_1) = (0^\circ, 79^\circ)$, whereas the wall reflection arrives from $(\varphi_2, \vartheta_2) = (81^\circ, 72^\circ)$. The time domain results are plotted where time is mapped into distance from the origin. The two spheres indicate the TOAs and the three color grades indicate three specific values of the results $|w_s(\varphi, \vartheta, t)|$, namely 0 dB, -12 dB, and -21 dB.

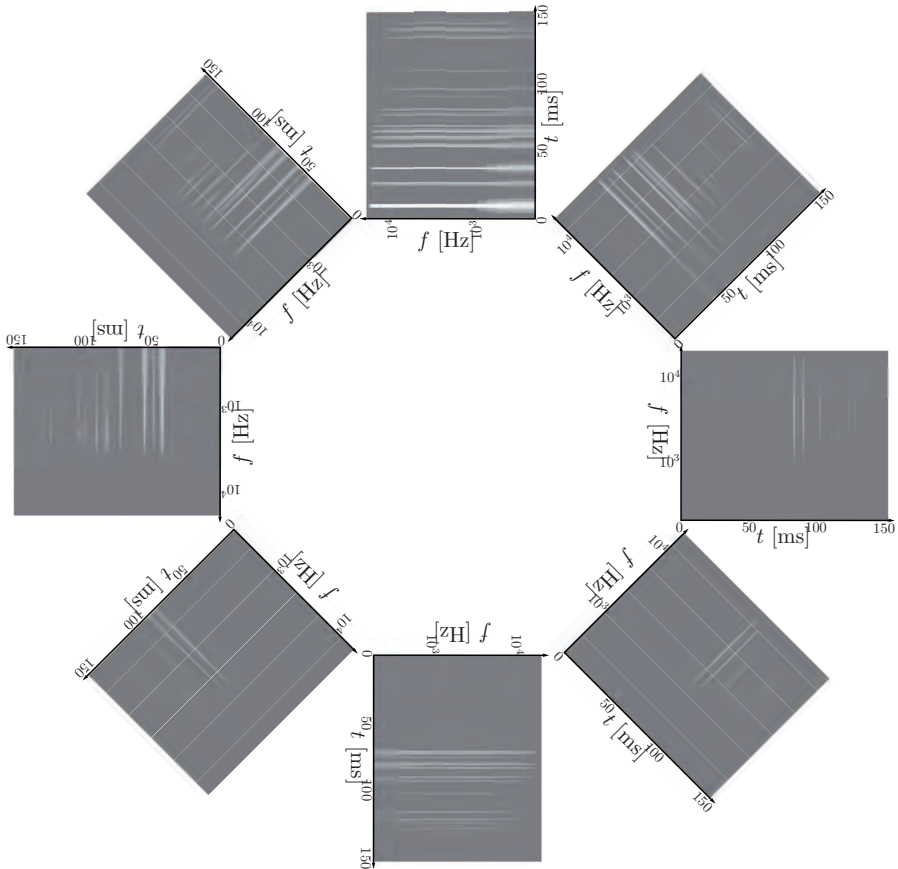


Figure 6.10: Three-axes multi impulse response representation with spatial arrangement of impulse responses corresponding to their angle of incidence. The impulse responses are plotted as time versus frequency representation. In the upper direction the direct sound can be noticed.

be used frequently in the interaction process because in most cases a higher spatial resolution is not required in the editing process.

6.3 Static Interaction

In this section the static interaction (SI) will be discussed. SI comprises the process of modifying the complete room impression without taking the reproduction system and the source parameter into account. SI can be interpreted as a further development of the current user interfaces for reverberation units based on convolution. The developments address the following two main properties of actual user interfaces:

- A single visualized impulse response measured or simulated with an omnidirectional microphone represents only an overall room impression. The impulse responses are not visualized in a direction-dependent way (e.g., (Altverb, 2009)).
- Several signal processing methods are applied to the impulse responses or to the output of the processing. The processing has to be specified by numerical parameter. For these, separate user interfaces (e.g., Buttons, Sliders) are used (e.g.,(Reverb, 2009)). This is not very intuitive.

These are addressed by the following new approaches:

- The visualization is extended to a direction-dependent impulse response representation.
- The processing is specified by a direct interaction of the user with visualization. New methods are developed to modify the properties directly in the visualization, without using a second interface representing the parameters for the modification.

First the basic principle for direct interaction using shaping surfaces is introduced. Then principles for the modification of single impulse responses are presented such as the inverse energy decay curve, time-variant filtering and decorrelation. These principles are extended to direction-dependent impulse responses afterwards.

6.3.1 Shaping Surfaces

In the previous section, it has been shown how a RIR can be analyzed and visualized using three-axes representations. In this section the modification of such a graphical representation is addressed. An intuitive way for the user to modify a graphical representation is a direct modification of visualizations using a dedicated tool. The problem with using this approach for an impulse response is the fine structure of such data, which has to be preserved to ensure a high auralization quality. Compared to the internal structure, modifications by the user should be relatively coarse. As a consequence, the modifications have to be in different resolution than the data used for the signal processing. For this reason, parametric surfaces are used in this work. The definition of such surfaces can be made using a limited number of control points, while the values of the resulting parametric surface can be evaluated at arbitrary resolutions. An example can be found in Figure 6.23. To modify a data representation, a parametric surface is generated above the visualization. Using this concept, the data are not required to be visualized with the parametric surface in order to have lower computational complexity. Furthermore, the user is always aware of the modifications which are represented by the parametric surface. In order to change the parametric surface, the user can modify the control points of the parametric surface. By this action, a modification surface is generated. The values for the signal processing are generated by sampling the parametric surface. The following steps are carried out during the interaction process:

1. Visualization of the data using an appropriate representation denoted by the transformation $\{\cdot\}$.
2. Generation of a parametric surface, spatially separated from the data representation, using the required number of control points and parameters specified by the user interface.
3. Modification of the parametric surface according to the desired modification of the visualized data.
4. Evaluation of the parametric surface at a resolution required for the signal processing.
5. Combination of the parametric surface with the data representation by an arbitrary mathematic operation.
6. Inverse transform of the data denoted by $\{\cdot\}^{-1}$.

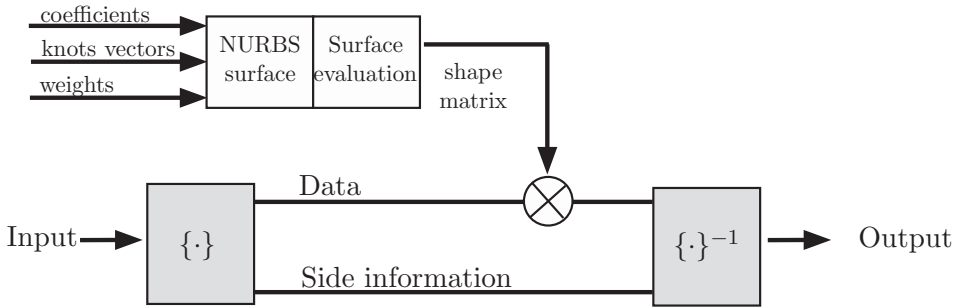


Figure 6.11: Block diagram for the concept of surface shaping

This process is used to realize modification of single impulse responses (e.g., time-variant filtering) as well as direction-dependent modifications. The preferred parametric surfaces within this study are Non-Uniform Rational B-Splines, as described in Section 2.6. These parametric descriptions can be evaluated and drawn according to algorithms presented in (Piegl and Tiller, 1997).

6.3.2 Interaction with Single Impulse Responses

The basis for the direct interaction with multi-impulse responses is the interaction with a single impulse response. Depending on the axis representation according to Section 6.2.1, one can consider the following main editing capabilities:

- Direct modification of the envelope applied to time domain data.
- Inverse Energy Decay Curve.
- Time-variant filtering.
- Frequency-dependent decorrelation.

The first is a common tool and will not be addressed further in this thesis. Applications can be found in commercially available products as given in (Reverb, 2009). The last three editing techniques will be described in the subsequent sections.

Inverse Energy Decay Curve

The energy decay curve (EDC) of a RIR plotted with a logarithmic amplitude scale is highly correlated with the perception of the decay of a diffuse sound field. Detailed fluctuations as shown in a time plot of a RIR are not perceived in detail. For this reason, a direct modification of the energy decay curve is proposed for sound design applications. After the definition of the desired EDC by the user, an envelope can be generated to modify the impulse response in a way that the new curve is obtained. The EDC $s(t)$ of an impulse response $h_1(t)$ can be calculated using the Schroeder integral, Eq. (6.3). In the case of a sampled room impulse response $h_1(n)$ with N samples, the following expression is used:

$$s_1(m) = T \sum_{n=m}^N h_1^2(n), \quad (6.5)$$

with T sampling interval and N representing the length of the RIR. Modifications by a user result in the energy decay curve $s_2(n)$ as shown in Figure 6.12 (a). To calculate the required envelope directly from the two curves $s_1(n)$ and $s_2(n)$, their difference is required,

$$s_d(n) = s_1(n) - s_2(n). \quad (6.6)$$

From the result the required envelope $e(n)$ is calculated with

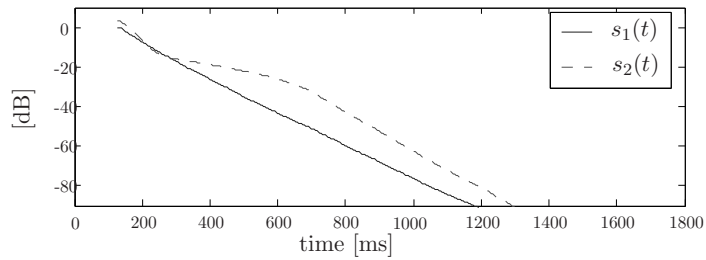
$$e(n) = \sqrt{1 - \frac{s_d(n) - s_d(n+1)}{h_1^2(n)T}}. \quad (6.7)$$

The envelope $e(n)$ generated using $s_1(n)$ and $s_2(n)$ is shown in Figure 6.12 (b). Eq. (6.7) represents the required envelope for the modification of the impulse response $h_1(n)$ to get the desired RIR $h_2(n)$ using

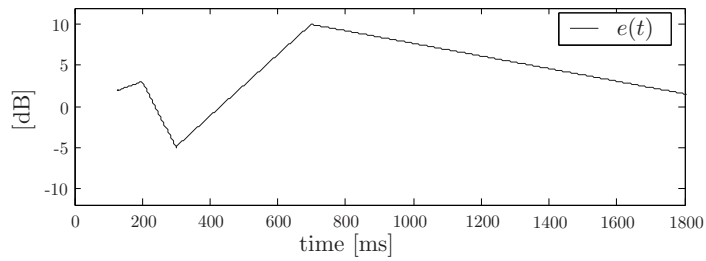
$$h_2(n) = e(n) \cdot h_1(n) \quad (6.8)$$

Time-variant Filtering

As described in Section 2.5 the discrete short time Fourier transform of a discrete signal given by $DSTFT \{p(n)\}$ is an appropriate basis for time-variant



(a) Energy decay curves



(b) Envelope

Figure 6.12: Direct modification of energy decay curves and corresponding envelope. $S_1(t)$ representation the original EDC and $s_2(t)$ is the version generated by the user. The generated envelope to modify the RIR is given by $e(t)$.

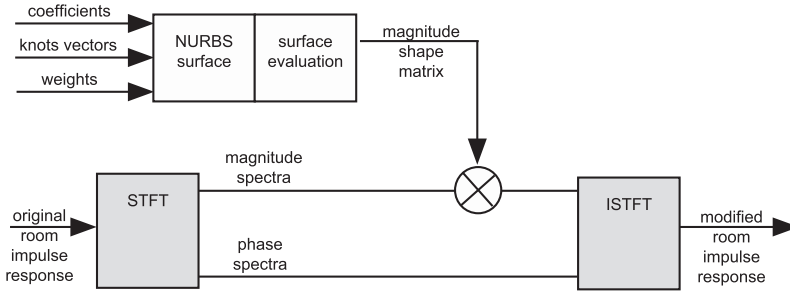


Figure 6.13: Block diagram for the proposed shaping of a room impulse response.

modification of signals if the parameters of the transformation are chosen appropriately. In the context of time-variant filtering, the user modifies the magnitude spectrum and the processing should avoid any time aliasing artifacts. To meet this requirement, first the analysis window w_a and the hop-size R_a as defined in Section 2.5 should be chosen such that Eq. (2.56) is satisfied. Second, the size of the DFT for each frame should be $k \geq (N_a + N_b) - 1$ where N_a and N_b are the number of windowed samples of the RIR and the sampled surface, respectively. This gives the anti-aliasing condition for the filtering process. Figure 6.13 shows a block diagram of adaptation of the shaping surface concept applied to time-variant filtering. Given a time-variant complex discrete spectrum $P(s, k)$ of the discrete impulse response or signal $p(n)$, where s represents the time index and k the frequency index in the STFT-domain, $P(s, k)$ can be separated into a magnitude and phase spectrum as

$$A(s, k) = 20 \log_{10}(|P(s, k)|), \quad (6.9)$$

$$\Theta_A(s, k) = \angle P(s, k). \quad (6.10)$$

While the phase spectrum $\Theta_A(s, k)$ is not modified, the user modifies a normalized NURBS-surface, which is evaluated in the same resolution (s, k) as the magnitude data $A(s, k)$. It is assumed that the NURBS representation is given in a logarithmic magnitude scale with 0 dB in its default setting. The evaluated and sampled NURBS-surface $S(u, v)$ is used for shaping the magnitude values of the room impulse response, which results in the new magnitude spectrum $B(s, k)$. Hence

$$B(s, k) = 10^{\frac{A(s, k) + S(u, v)}{20}}, \quad (6.11)$$

with

$$\begin{aligned} u &= [0, 1/s, 2/s, \dots 1] \\ v &= [0, 1/k, 2/k, \dots 1]. \end{aligned} \quad (6.12)$$

In this way, arbitrary time-variant modifications are combined with the advantage of an intuitive interaction for the user. The processing is closely related to the cross synthesis of two sounds described by (Serra, 1989), while the room impulse response represents one and the evaluated NURBS-surface represents the other sound. The complex spectra $P'(s, k)$ for the shaped room impulse response are made up of the phase spectra $\Theta(s, k)$ and the modified magnitude spectra $B(s, k)$:

$$P'(s, k) = B(s, k) \cos(\Theta(s, k)) + iB(s, k) \sin(\Theta(s, k)). \quad (6.13)$$

Now the IDSTFT of $P'(s, k)$ is computed using Eq.(2.57). In this case the synthesis window can be neglected because the phase response is kept

$$p(n) = \sum_{s=-\infty}^{\infty} p_s(n - sR_s) \quad (6.14)$$

with

$$p_s(n) = \frac{1}{N} \sum_{k=0}^{N-1} \left[e^{-i\frac{2\pi}{N}sR_s k} P(n, k) \right] e^{-i\frac{2\pi}{N}nk}, \quad (6.15)$$

where R_s denoted the synthesis hop-size. The NURBS-surface used for modifications, the DSTFT representation and the EDR plot are shown in Figure 6.14. During the reverberation phase, the amplitude is first reduced and then amplified during the late decay. The STFT provides a good view on the processing of the discrete impulse response. The modification of the NURBS-surface is reflected clearly. The energy decay relief plot delivers a very good estimation of the perception of the room effect after the convolution process. It is important to note that a strong amplification in the late part of the decay can lead to a strong enhancement of measurement noise. To avoid this, the impulse response should be sampled with a very high signal-to-noise ratio (SNR). An alternative is the reconstruction of the late part using filtered random noise.

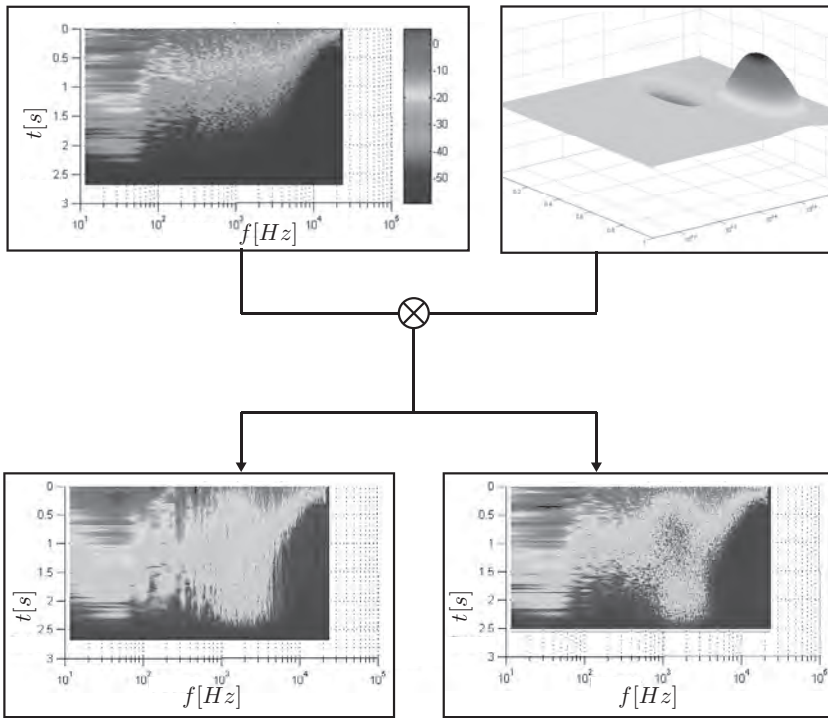


Figure 6.14: Example for a time variant shaping of an impulse response using a shaping surface. The original impulse response (top left) is given as well as the NURBS-surface (top right) used for shaping. The result is shown as STFT-plot (bottom right) and EDR-plot (bottom left).

6.3.3 Interaction with Direction-dependent Impulse Responses

The interaction with direction-dependent impulse responses requires the selection of the desired direction for editing before applying the methods given in the previous section. The extension required for the integration in a user interface is a time and angle selection process, corresponding to a *select and modify* paradigm. The following steps are performed in the interaction loop by the user:

1. Choose the desired *high resolution data*
2. Select time range and appropriate temporal window
3. Select angle range and appropriate spatial window

4. Choose a representation of the resulting impulse response
5. Modify the selection with the tool given in Section 6.3.2
6. Process the data under consideration over the chosen time and angle window

When using a representation with a limited number of directions (see Figure 6.10), one of the visualizations can be selected directly (e.g., by clicking on the desired representation). Due to the fact that the interaction should be independent of the reproduction system, the underlying data always has a high spatial resolution, even if it is not visualized in the user interface. The resolution has to be adapted to the visualization. After the interaction, the parameters specified by the user have to be applied to the appropriate subset of the high-resolution data. The high-resolution data is given by the weights $w(\varphi, \vartheta, k)$, which are the result of a plane wave decomposition. It is assumed that the required weights according to the grid of the PWD have been applied. In this section, impulse responses are studied in the time domain. After applying a temporal Fourier transform, they are given by $w(\varphi, \vartheta, t)$. In this section, a spatially discrete representation is considered. The resulting impulse representing the pressure $p(\mathbf{x}, t)$ in the center \mathbf{x} of the sphere can be calculated using

$$p(\mathbf{x}, t) = \sum_{i=0}^N w_i(\varphi_i, \vartheta_i, t), \quad (6.16)$$

where N represents the number of grid points. A separable grid is common for the directional weights (e.g., Chebyshev). In this case Eq. (6.16) can be represented using two sums representing the azimuth and co-elevation directions

$$w(\mathbf{x}, t) = \sum_{i=0}^N \sum_{j=0}^M w_i(\varphi_i, \vartheta_j, t). \quad (6.17)$$

To extract a subset of data $p_s(\mathbf{x}, t)$, a discrete window function for the azimuth and co-elevation is applied. The window functions are denoted by $win_s(\varphi_i)$ and $win_s(\vartheta_j)$, respectively:

$$p_s(\mathbf{x}, t) = \sum_{i=0}^N \sum_{j=0}^M win_s(\varphi_i) win_s(\vartheta_j) w_i(\varphi_i, \vartheta_j, t). \quad (6.18)$$

The reduction of resolution used in the previous section was generated by applying Eq. (6.22). This equation can be extended by a temporal window $win_s(t)$ for the three-dimensional selection in a PWD, hence

$$p_s(\mathbf{x}, t) = win_s(t) \sum_{i=0}^N \sum_{j=0}^M win_s(\varphi_i) win_s(\vartheta_j) w_i(\varphi_i, \vartheta_j, t). \quad (6.19)$$

During the interaction process, the user can select an angle range and a temporal range to specify the area that he wants to modify. An appropriate window function is applied or can also be specified by the user. The result $p_s(\mathbf{x}, t)$ is visualized using the representation given in Section 6.2. A direct interaction method is applied to the selection. For the processing the same window function is applied to the parameter specified by the user. Subsequently, the processing is applied to the high-resolution directional weights. Several quantities can be derived from room impulse responses. The most relevant ones are specified in (ISO 3382, 2000). Such quantities are related to specific more or less perceptual attributes. These quantities can be modified by the user using an automated spatio-temporal selection. This is achieved by using dedicated macros for the spatio-temporal window. As example, the acoustic parameter Lateral Energy Fraction (LF) is considered. The value can be calculated from the room impulse response using

$$LF = \frac{\int_{0.005s}^{0.08s} p_8^2(t) dt}{\int_{0.005s}^{0.08s} p^2(t) dt}, \quad (6.20)$$

where $p(t)$ and $p_8(t)$ denote the impulse responses measured with an omnidirectional microphone and a figure-of-eight microphone, respectively. The desired impulse responses can be calculated from the plane wave decomposition by

$$p(t) = \sum_{i=0}^N \sum_{j=0}^M w_i(\varphi_i, \vartheta_j, t). \quad (6.21)$$

and

$$p_8(t) = \sum_{i=0}^N \sum_{j=0}^M win_s(\varphi_i) win_s(\vartheta_j) w_i(\varphi_i, \vartheta_j, t), \quad (6.22)$$

using the window functions

$$win_s(\varphi_i) = \cos(\varphi_i) \quad (6.23)$$

and

$$win_s(\vartheta_j) = \cos(\vartheta_j), \quad (6.24)$$

where $\varphi_i \in [0, 2\pi]$ denote the sampled azimuth angle and $\vartheta_j \in [0, \pi]$ the sampled co-elevation. If a gain is applied to the directional weights which are windowed in this way, the above parameter is modified. The same principle can be used with other acoustic parameters. In the editing considered so far, the user selects a range of impulse responses in terms of angle and time, which are treated as a single one, using the interaction methods for single impulse responses discussed in Section 6.3.2. For special parameters a direct manipulation in a direction-dependent way is useful. This includes gain of the directional weights. One can consider enhancing the reverberation from a specific direction for artistic sound design purposes. A solution to edit quickly the direction-dependent parameter, is the application of a spatial envelope. This is the application of the shaping surface concept in a spatial arrangement. In the case of 2D reproduction, a surface in the form of a disk can be used, representing the desired value in the height, the direction in the angle and another parameter (e.g., time) as the radius. An example of a three-dimensional spatial envelope will be given in Section 6.5.2 in Figure 6.24(b). A parametric line linked to a polar plot representing the direction as an angle and the value as the radius is also possible. Such representations will be used also for the directivity of sources as given in Figure 6.18.

6.4 Dynamic Interaction

The dynamic interaction process denotes the part of interaction which is performed in real time during the auralization. For this reason, it is an integral part of the reproduction-system dependent rendering. In this section only the user interfaces and general signal processing concepts will be discussed. The reproduction-system dependent part of the signal processing is discussed in Chapter 7.

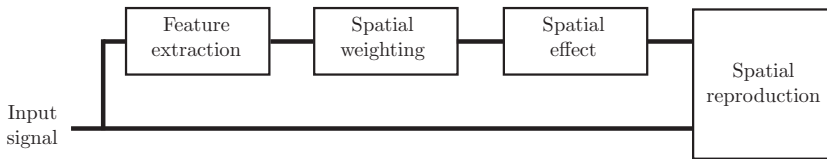


Figure 6.15: Block diagram of spatial processing dependent on the source signal. The extracted feature can be a mean level which is weighted by a spatial envelope. The side chain signal controls the value of a specific parameter in a spatial effect system. A controllable parameter can be the direct-to-reverb ratio.

6.4.1 Source Signal

The source signal itself can be used to modify the signal processing. For this reason, features are extracted from its audio stream. Such features are weighted in a direction-dependent way by applying a spatial envelope. The result can be used to modify parameters, e.g. level of the impulse responses (see Figure 6.15). To illustrate the processing, a specific example will be given. As a particular feature of a spatial audio signal the direction-dependent mean level can be considered. This can act as a control value for the direction-dependent control of the direct-to-room ratio. If this extracted feature is inverted, the room simulated will be enhanced in the directions in which low values of the direct sound occur. Furthermore, the spatial average of the value can be used to compress the dynamic. One can consider applying a dynamic processor on the signal feed to the room simulation. This dynamic processor extracts the level of the audio stream and uses it to modify the dynamics. The direct signal is reproduced directly using a spatial reproduction system. The signal used for the room simulation has a limited dynamic range. The direct-to-room ratio will be inversely proportional to the level of the direct signal. As a result, the room signal is emphasized in phases of low level of the source signal and such directions. To complete the example, the room effect of a low level source signal results in a strong room effect from the opposite direction.

6.4.2 Source Position

Two-dimensional positioning In case of a two-dimensional reproduction system, an intuitive method of source positioning is the movement of a graphical representation on a two-dimensional plane. This method was used in early

spatial sound design user interfaces for WFS (Melchior et al., 2003), which was more recently presented in (Meltzer et al., 2008). Beside the real time movement the recording of the movement is required. This process is called *automation*. From such a real-time recording of movement (i.e., sampling of a movement of the user), parameterized curves can be derived using methods of interpolation or approximation. The reader is referred to specific literature (e.g., (Piegl and Tiller, 1997; Boor, 1994)) for a detailed discussion. If the timing and the motion path need to be edited separately, a visual connection between these two representations is required. A solution is the use of parametric curves to describe the movement, which also can be edited by the user. The motion path is enhanced with markers to represent a spatial position in time. A possible realization is the use of Bézier curves for the motion path. An n th-degree Bézier curve is defined by (Piegl and Tiller, 1997)

$$C(u) = \sum_{i=0}^n B_{i,n}(u)P_i, \quad (6.25)$$

where $0 \leq u \leq 1$ defines the position on the curve. The basis functions are n th-degree Bernstein polynomials given by (Piegl and Tiller, 1997)

$$B_{i,n}(u) = \frac{n!}{i!(n-i)!} u^i (1-u)^{n-i}. \quad (6.26)$$

The geometric coefficients P_i are called control points and can be used to edit the curve. To define a movement, a second function $u = f(t)$ is required to map the values of u defined by χ_1 to χ_4 to specific times τ_1 to τ_4 (see Figure 6.16). It is intuitive to allow the editing of the points directly on the curve. In this way points in space can be associated with positions in time. The example in Figure 6.16(b) shows different velocities of the source for the different sections on the motion path. The section represented by $u \in [0, 0.5]$ is moved in a short time segment $t \in [\tau_1, \tau_2]$ while the remaining part of the path $u \in (0.5, 1]$ is moved with different slower velocity as in time segments $t \in (\tau_2, \tau_4]$. This approach can be used with different parametric curves as well as with direct recording of motion paths by the user and with a subsequent conversion of a sampled movement to a parametric representation. The two representations can be connected in an intuitive way by using the third dimension. This is depicted in Figure 6.17, which includes the third dimension as a time axis. This approach also enables the visual connection between single coordinates versus time representation. In common interfaces of digital audio workstations

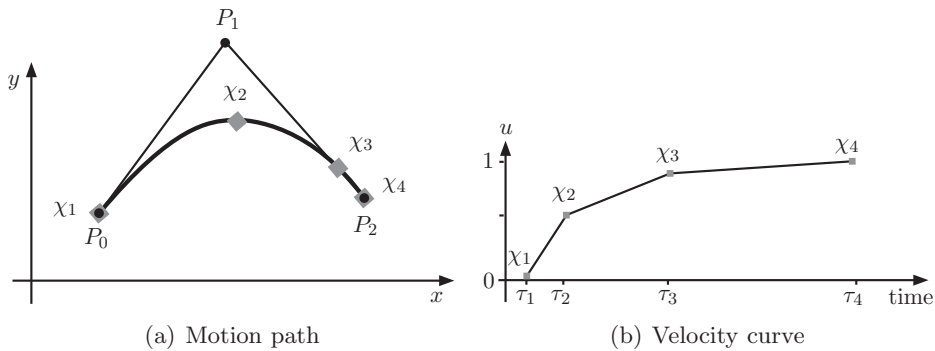


Figure 6.16: (a) Motion path of a sound source in the xy -plane. χ_1 to χ_4 represent the time markers, connecting the motion line to different times τ_1 to τ_4 . (b) corresponding velocity curve including the markers given in (a).

(DAWs) only one-dimensional data are associated with a time axis. In case of movement, only the values of the coordinate system used are edited by the user.

Three-dimensional positioning Time markers and parametric curves can also be used for 3D positioning and automation of sound sources. The previously described extension including the time axis is not possible in this case, because all three dimension are required to show the source position. Furthermore, a 3D positioning device is strongly recommended, to record a movement in real time. Without such a device the different dimensions have to be recorded separately, which is not very easy. The user interface concept given in section 6.5 will support 3D scenarios. Another solution is to edit the motion path on a user interface utilizing different views. Afterwards the timing can be recorded in realtime using an interface which represents the actual position on the path (e.g., a fader). Such an approach has been evaluated during this project and was well received by potential users (Schneidmadel and Seidenzahl, 2005).

6.4.3 Source Directivity

Editing source directivity is challenging because a four-dimensional dataset has to be edited. The directivity depends on the 3D orientation of a source and is frequency-dependent. A solution, as applied in state-of-the-art software,

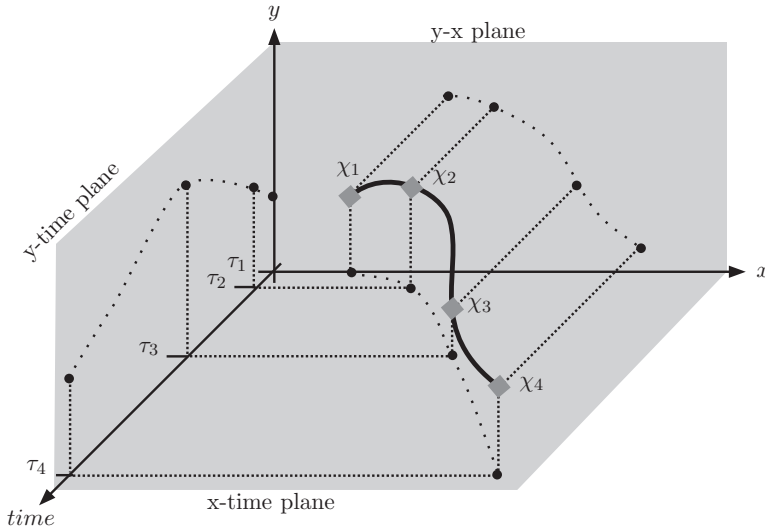


Figure 6.17: (a) Motion path of a sound source including the timing in a 3D visualization. χ_1 to χ_4 represent the time markers, connecting the motion line to different times τ_1 to τ_4 .

is the separation into two intersecting planes describing the vertical and the horizontal directivity. Such representations are plotted as polar diagrams. The concept of spatial envelope and shaping surfaces can be applied successfully to the modification of source directivity. Parts of a user interface prototype for source directivity editing developed throughout this thesis (Scheck, 2008) are shown in Figure 6.18. The user can modify the directivity of a given plane by editing the polar diagram directly. The frequency can be selected separately. Based on these edits, a three-dimensional directivity is interpolated. The polar diagram can be modified using the marker on the curve. This method is a simple implementation of the spatial envelope described in Section 6.3.3.

6.4.4 Signal Processing Depending on Source Parameter

Based on the changes of source properties, the auralized acoustics of a room should also change. While the diffuse part can be considered as independent of the source parameter, the early part of the impulse response depends on the source directivity and position. (Hulsebos, 2004) proposed the pseudo-extrapolation of extracted reflections based on the assumption of first-order specular reflections. The work of (Pellegrini, 2001) supports the proposition

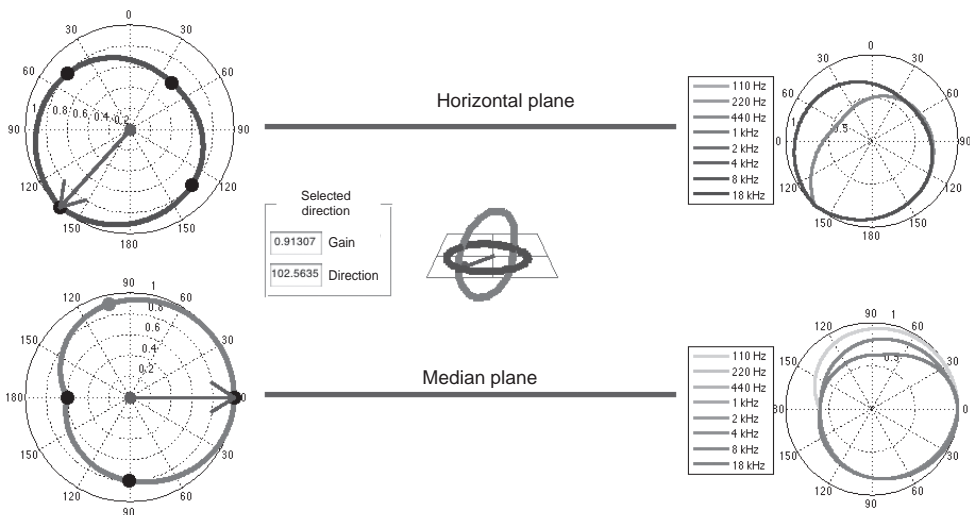


Figure 6.18: User interface for directivity manipulation based on direct interaction with a polar plot. The control points of the splines in the left polar plots can be modified to achieve the desired directivity for the selected frequency.

that, for the perception of distance and room size, not the detailed reflection pattern is required, but a limited number of strong reflections is sufficient to stimulate the perception. The experiments presented in Chapter 8 also support this hypothesis. For this reason, the pseudo-extrapolation will be used to reflect the changes of source parameters (i.e., position, direction and directivity). The term pseudo-extrapolation refers to the fact that the exact extrapolation based on the available data is not possible, but that changes of reflections are calculated. This can be seen as a perceptually motivated solution. For a physically motivated auralization, a measurement for each source position, directivity and orientation would be required.

Pseudo-extrapolation of Early Reflections

The mirror image source model served as a modeling tool for rooms of specific geometries (Cremer and Müller, 1976) (Vorländer, 2008). It was extended to arbitrary geometries by Allen and Berkley (1979) and Borish (1984). For the pseudo-extrapolation, an infinite plane located in the middle between the initial source position and a detected reflection is assumed. Referring to the geometry in Figure 6.19 and using the original source position \mathbf{s}_1 and the measured reflection \mathbf{r}_1 , the plane A is constructed using a normal vector \mathbf{n} given by

$$\mathbf{n} = \frac{\mathbf{r}_1 - \mathbf{s}_1}{|\mathbf{r}_1 - \mathbf{s}_1|} \quad (6.27)$$

and the foot point \mathbf{a} defined by

$$\mathbf{a} = \mathbf{s}_1 + \frac{1}{2}(\mathbf{r}_1 - \mathbf{s}_1). \quad (6.28)$$

Using the constructed plane, a new reflection position \mathbf{r}_2 for a new source position \mathbf{s}_2 is calculated with (Vorländer, 2008)

$$\mathbf{r}_2 = \mathbf{s}_2 - 2d\mathbf{n} \quad (6.29)$$

with

$$d = (\mathbf{s}_2 - \mathbf{a}) \cdot \mathbf{n}. \quad (6.30)$$

In addition to the new position, a new gain of the reflection is required. Under the assumption that the original gain of a reflection is available from the analysis process, only the change has to be taken into account. The new source position, based on the geometrical process described above, results in a

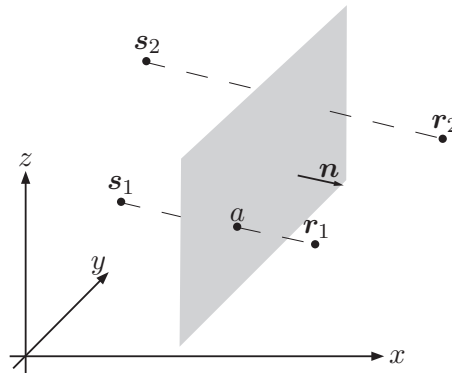


Figure 6.19: Geometry used for the pseudo-extrapolation. A reflection r_1 for a given source position s_1 is extrapolated to the new reflection position r_2 for a new source position s_2 . For the geometrical calculation a surface characterized by the normal vector n is used.

new distance. The reflections are assumed to be generated by point sources. As a result, the $1/r$ rule can be applied and the amplification or attenuation for the new position can be calculated.

6.5 User Interfaces for Spatial Sound Design

In this section two concepts for user interfaces are proposed. Both are designed for integration into a desktop studio environment. The first is based on a MATLAB graphical user interface. The second interface makes use of novel interaction devices and is designed for two-handed interaction. First the results of early experiments with the novel user interfaces are given. These lead to the development of the two interfaces discussed in the next sections. In the context of this work, only studio or mixing environments, will be considered. The reader is referred to (Melchior et al., 2006b) for a sound positioning user interface dedicated to live performances (e.g., theater or opera production), which was designed by the author in cooperation with the Bregenz Festival.

Early prototypes A method investigated by the author is the use of Augmented Reality (AR) technology for source positioning (Melchior et al., 2005). The reader is referred to (Azuma, 1997) for an overview and introduction to AR. The methods of AR were used to display or augment the position of

sound sources in the view of the user. For this purpose, a head-mounted display was used. A helmet was equipped with two video cameras generating a video image of the environment in the view of the user. A high-end graphic PC was used to extract the position of the user and augment graphical objects in the video images. Further investigation of the hardware and software used as well as the perception of audiovisual representations can be found in (Melchior et al., 2006a). Two setups have been studied. The exocentric setup (Figure 6.20(d)) augments the sources in the real scene at the position where they are audible. The user can modify the sources using a tablet. In the egocentric setup (Figure 6.20(b)) the user interacts with a small model on a table top, as shown in figure 6.20(d). In the field of user interfaces for AR and virtual reality (VR), such representation is also called the *world in miniature* metaphor (Stoakley et al., 1995). For the use in a real mixing environment the egocentric setup has to be extended to enable the positioning of sources outside the reach of the user (e.g., sources outside the loudspeaker setup). This is easily possible with the exocentric setup. Furthermore, it can be used to position sources in a three-dimensional space. For the application of AR technology in actual production environments, the use of head mounted displays (HMD) is not appropriate because they limit the mobility of the user and can be exhausting over long time use. Further development in AR will improve the situation, but wearing special devices is critical in sound design environments. For this reason, a new concept for the user interface was developed as given in in Section 6.5.2. No special devices have to be worn and it uses the exocentric set-up.

Current interfaces for source positioning Still state-of-the-art is the positioning of sources using a mouse or pen on an appropriate display. The use of a pen display has proved to be useful, because the required accuracy for position can be combined with a flexible interface. A drawback is the fact that such an interface cannot be used without visual feedback. For this reason a hardware mixing console is often preferred to a software simulation (Albrecht, 2005). The design of dedicated hardware user interfaces is beyond of the scope of this thesis, but will be an important part of future developments in spatial sound design tools. In addition to the simple positioning of sound sources, one can think of several extensions of this approach. The concept of different layers and hierarchies of an auditory scene is described in (Melchior et al., 2003). The user can arrange the source in groups in order to move all members of such a group simultaneously. Furthermore, such groups and sources



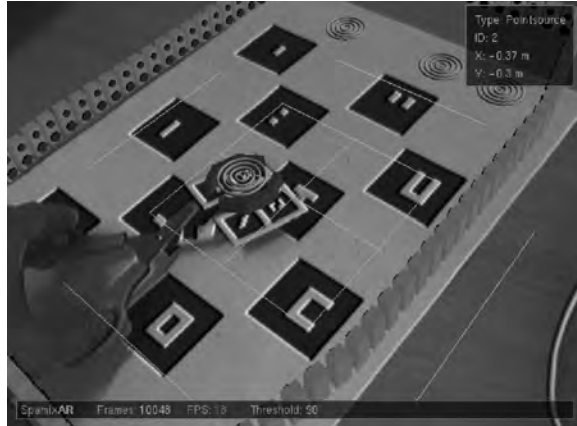
(a) Working user.



(b) Screen shot of the HMD (View of the user).



(c) Working user.



(d) Screen shot of the HMD (View of the user).

Figure 6.20: Augmented reality source positioning system. Egocentric setup (a) and (b) . Exocentric setup (c) and (d).

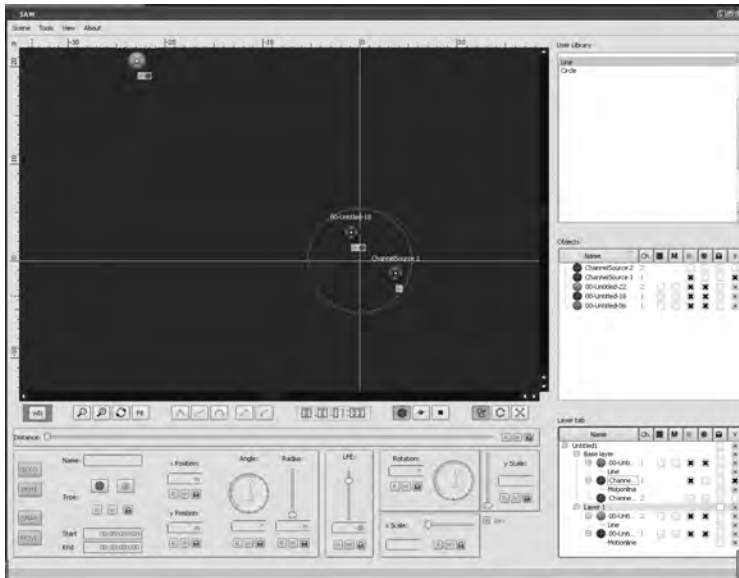


Figure 6.21: Typical 2D authoring tool for spatial audio (WFS). The main windows presents a top view of a spatial audio scene. In the lower section the source specific parameter can be edited. The sound sources can be structured on different layers and groups in the lower right side. Above an overview of all sources can be found.

can be arranged in layers to combine sources that belong together in terms of music, sound effects or dialog. These concepts have now been implemented in the spatial audio workstation (SAW) shown in Figure 6.21. A interesting extension of a graphic user interface using standard elements is the zooming interface paradigm (Raskin, 2000). This approach uses the graphical representation not only for positioning but also to display several parameter if the zoom is adequate. The use of a zooming interface was studied in detail in (Schneidmadel and Seidenzahl, 2005). Because of the extensive graphical programming required, such a system has not been realized until now.

6.5.1 Standard Interface for Spatial Sound Design

The standard user interface for static interaction was realized using the MATLAB processing and graphical user interface (GUI) environment. The processing and the user interface are reduced to two-dimensional processing. This

is achieved by summing the directional weights of a PWD for all co-elevation angles as described in Section 6.3.3. The graphical user interface is a pure desktop application. The user interacts using a mouse and a keyboard. The aim of this interface is to have a prototype implementation of the processing in order to get feedback from potential users. The MATLAB GUI has to be combined with an adequate auralization engine. Only the required reproduction data and high resolution data are generated by the interface. The reproduction data can be used by a rendering framework for wave field synthesis, or in combination with convolution engines. The following modules are realized:

1. Simulation module. Implementation of a mirror image source model and diffuse sound model to simulate ideal plane wave decomposition of a modeled room.
2. Editing module. Implementation of the static interaction described in the previous sections
3. Adaptation module for WFS as described in Section 7.2
4. Adaptation module for virtual microphones (VM) as described in section 7.3

A screen shot of the editing module, including the shaping surface interface, can be found in Figure 6.22. The graphical user interface uses a MATLAB toolbox for the signal processing developed in this thesis. This toolbox includes all the interaction methods described and can be used to extend the GUI in the future. Positioning of user interface elements was designed according to the interaction process of the basic layer. In the upper left corner the user can select an angle and time segment. By using a specified window function, the resulting impulse response for the selection is displayed on the upper right. The type of representation can be selected below the interface. In the main area on the lower right, modifications to the selected spatial segment can be made. Envelope editing, energy decay curve editing and shaping surface editing have been implemented. The complete result is visualized in the lower left hand side using a two-axis representation including directional information. The data can be exchanged as *high resolution data* and be further processed by the other modules. A complete description including the wave field synthesis reproduction can be found in (Melchior et al., 2010).

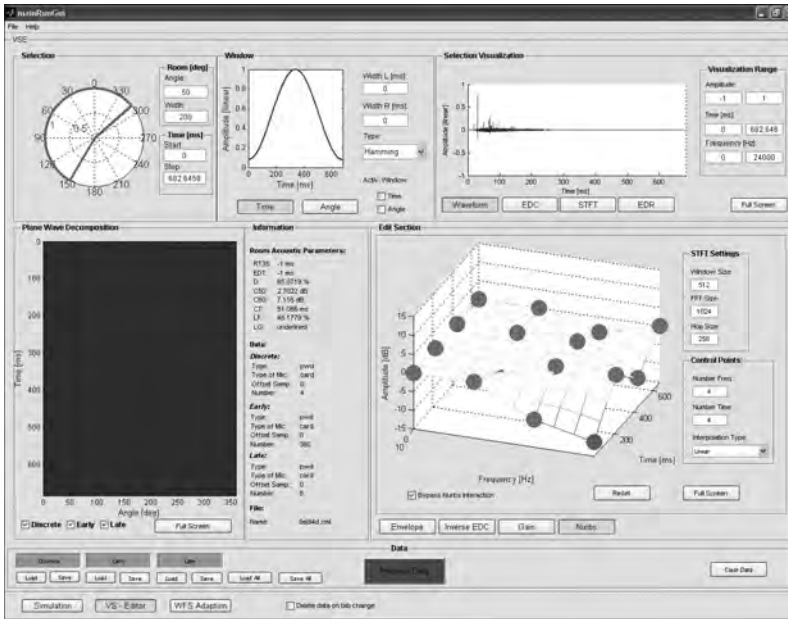


Figure 6.22: Standard user-interface based on MATLAB (Editing module). The interface is structured according to the steps of user interaction in a clockwise manner. First (upper left) an angle and time window can be selected. The extracted (spatio-temporal filtered) impulse response is shown in the upper right using selectable visualizations. In the lower right side, different interaction methods can be applied. The shaping surface is currently selected. On the lower left, the result is visualized using a multi-trace representation.

6.5.2 Extended Interface for Spatial Sound Design

The interface and visualization possibilities of a standard GUI are limited in the context of the requirements of 3D spatial sound design. Besides the interface based on standard user interface elements, parts of an extended user interface were realized using the AR system *Studierstube* (Schmalstieg et al., 2002). The system was interfaced to a MATLAB processing kernel used for the signal processing. Afterwards the visualization was performed by the *Studierstube* system using *Open Inventor* (SGI, 2008). The *Open Inventor* scene graph was extended with specific elements to enable the desired interaction and the control of a WFS-reproduction system.

Preliminary considerations The aim of the extended interface is to support the 3D visualization and interaction techniques in an optimal way. Most of the visualizations proposed in this work are three-dimensional. For this reason, appropriate techniques for direct manipulation of three-dimensional visualizations have to be used. The environment used is capable of supporting various display hardware including stereoscopic displays. To enable fast interaction, an asymmetric bimanual technique is proposed. This means that the user interacts with two hands simultaneously. The theoretical foundation behind the bimanual 3D interaction was given by (Bowman et al., 2005; Guiard, 1987). Using their principles to characterize the role of the hands, identifies the task of direct manipulation of impulse representations as an asymmetric task. This means that one hand is used for navigation and modifying the view, while the interaction is performed with the other hand. As a common example, handwriting on a blank sheet of paper can be considered. While the dominant hand is writing the non dominant hand is used to position the paper and fix its position. (Balakrishnan and Kurtenbach, 1999) showed that the two-handed interaction in a bimanual asymmetric task, similar to the desired one in the actual user interface, speeds up the completion time and is graded better in the subjective preference of the subjects. As a consequence the direct interaction is based on a system where the non-dominant hand is used to navigate and position the visualization and the dominant hand is used for the direct interaction.

Visualization Two key parts have to be represented by the user interface. First, the graphical representations as developed in the previous sections have

to be visualized. To facilitate the easy navigation and orientation of the user, the current reproduction system layout (i.e., loudspeaker placement) is also included. Secondly, graphical user interface elements to specify parameter and perform system control operations are also required. These are represented in two-dimensional user interface elements. As a result, displays for the 3D representation of the impulses response as well as for the 2D GUI are required. To edit the graphical object representing the spatial audio setup, or impulse response, the virtual 3D space is used. For 2D user interface elements for system control, a 2D interface on the pen display is used.

Navigation To enable an intuitive two-handed interaction, the *world in hand* metaphor (Houde, 1992) is used for the navigation. In this concept, an object (world) is attached to a navigation device which enables the rotation and translation of the object of interest to the needs of a interaction task. The *world in hand* metaphor seems to be optimal for viewing discrete, relatively compact data objects (Ware, 2004). This holds for the visualization of a 3D audio setup as well as for the impulse response visualization described earlier. To enable the two-handed interaction with the *world in hand* metaphor, a six-degree of freedom (DOF) interaction device is required. Another requirement as stated earlier is the integration of the system in a studio environment. For this reason, it should be possible to switch very quickly to different hardware user interfaces. Therefore, a device that can be placed on a desktop is desirable. A possible solution is a space mouse or space ball controller, which enables the 6-DOF manipulation of an object attached to its coordinate system.

Interaction The dominant hand of the user uses a pen for interaction. Such a pen can be used directly on the pen display for 2D interaction (Sutherland, 2003). When the user lifts the pen into the interaction volume, it is associated with a 3D cursor in the virtual space similar to the selection technique described by (Grossman and Balakrishnan, 2006). Using direct mapping of the real volume to the virtual volume, the pen acts like a mouse in a three-dimensional space. The pen is equipped with a button to enable selection tasks. If the pen is used directly on the pen display, the contact of the pen with the surface can also be used as click. Figure 6.23 presents a mock-up of the user interface. A detailed view of different realized interaction techniques can be found in the screenshot in Figure 6.24.

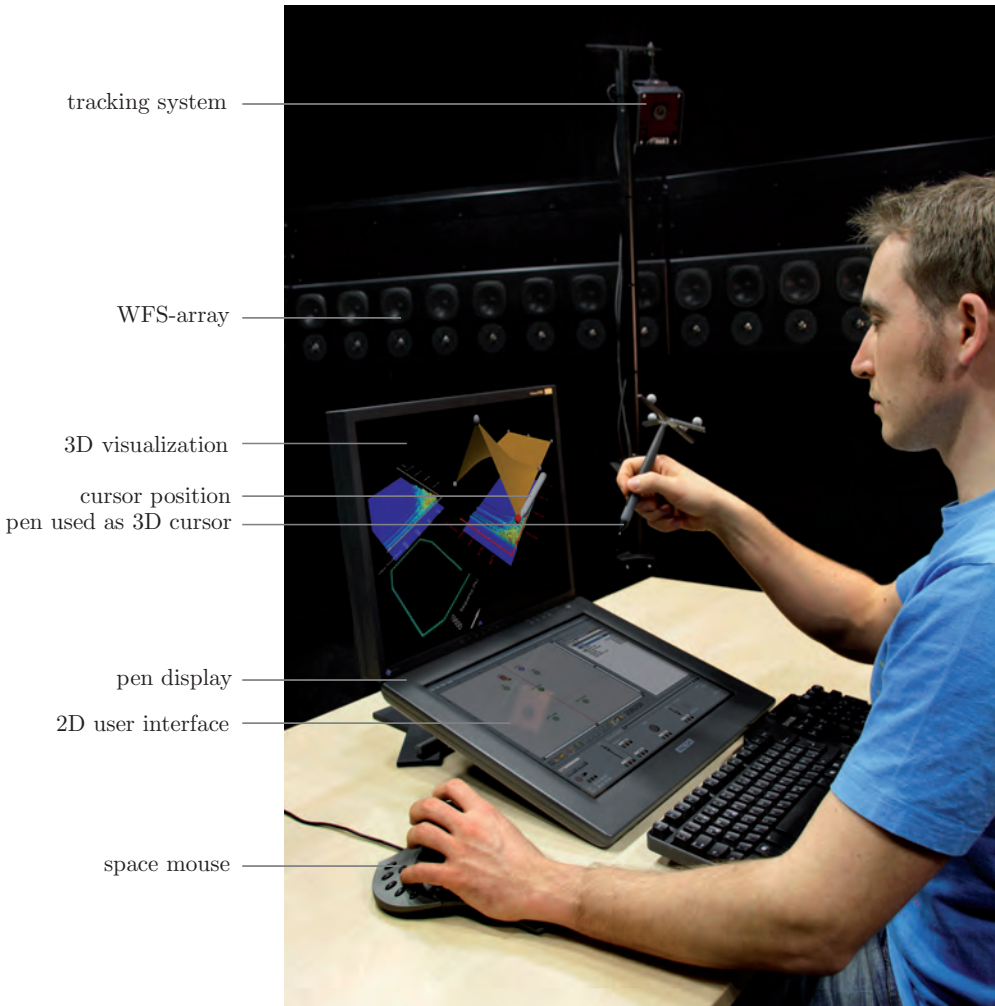


Figure 6.23: Proposed extended user-interface. The user can use a tracked pen to modify a 3D representation of the impulse response data. The pen can be moved in spaced in order to control its virtual representation. A space mouse is used to control the position and rotation of the 3D impulse response representation. A pen display is used for 2D user interface representation to control the system. This can be controlled by the pen when it is moved on top of the display. A detailed view of the screen shot of the system can be found in Figure 6.24.

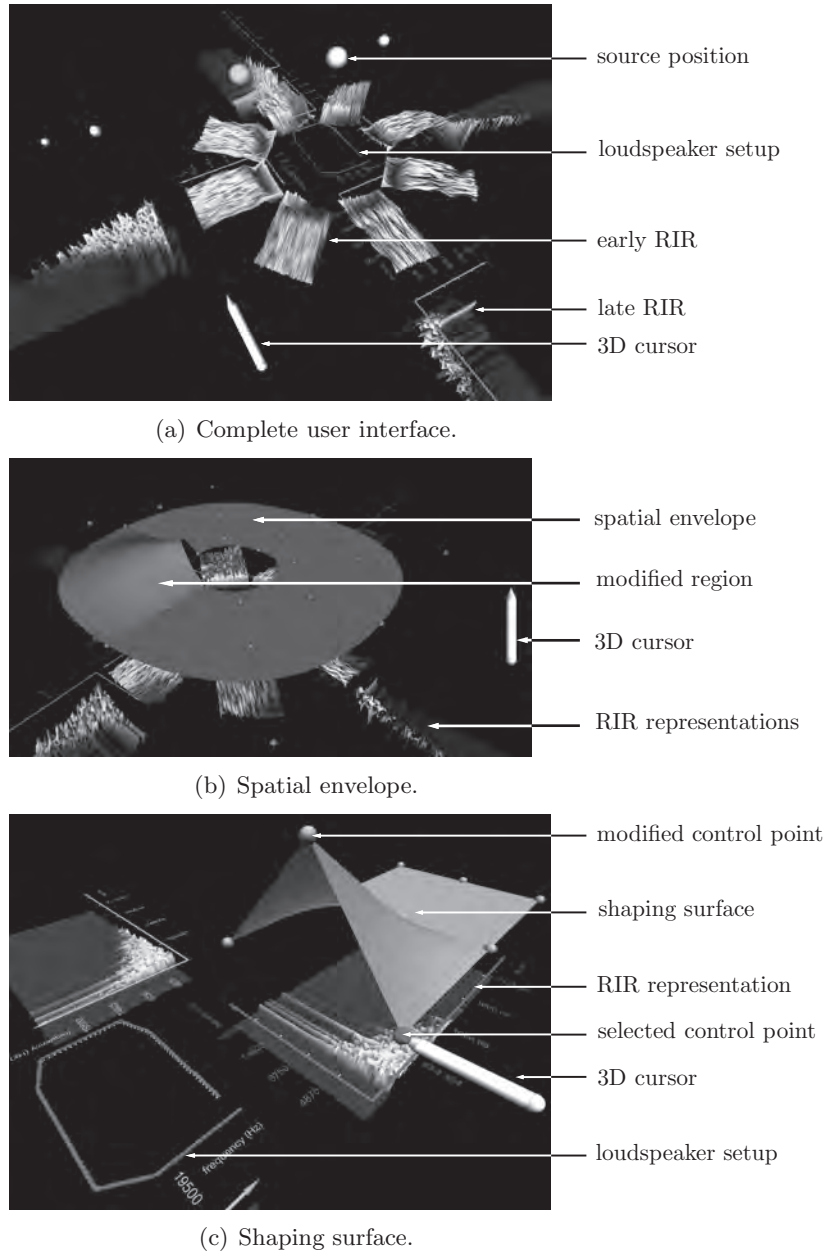


Figure 6.24: Screenshot of the extended user interface.

Prototype system To demonstrate the interaction method to potential users, some specific models are implemented. Emphasis was put on the concept of shaping surfaces. A screen shot of the extended user interface is given in Figure 6.24. A representation of eight directions for the early part of the impulse responses and four directions for the late part is shown in Figure 6.24(a). The prototype was used in conjunction with a WFS system. For this reason, the layout of the loudspeaker setup was also visualized in the middle. This was used as an orientation for the user. The virtual element representing the tracked pen of the user can be seen as well as different spheres representing the position of direct sound and reflections. The user can select a specific surface or source by moving the pen into its bounding box and clicking the button on the pen. In case of a source, the position can be modified while dragging the source to the desired position. In case of an impulse response representation, a shaping surface will be displayed. The detail of a shaping surface is shown in Figure 6.24(c). The spheres represent the control points of the NURBS surface shown. The point in the middle of the surface has already been shaped by the user. The virtual pen is used to select the right control point. Next, this point can be modified by moving the pen to the desired position. An example of spatial layer editing using the spatial envelope can be found in Figure 6.24(b). A specific direction has been emphasized by the modifications of the user.

6.6 Summary

In this chapter, the interaction with plane wave decomposition was studied. First a taxonomy of impulse response visualization methods and different diagram types was given. The different advantages for the use in a sound design process have been discussed. Based on these representations, the interaction process has been structured into a static and a dynamic interaction part. The static interaction denotes the spatial sound design process which is independent from the reproduction system. For the static interaction, novel representations of spatio-temporal impulse response representations have been developed. Based on this visualization, a new direct interaction method was developed. The novel method of shaping surfaces was applied to single and multi impulse response representations. The advantage of these new techniques is that a graphical representation of a direction-dependent impulse response can be modified in a direct interaction process by the user. The simultaneous editing of spatio-temporal data is possible within a single user interface. Due to the use

of known visualization techniques, these representations can directly be understood by a professional user. Furthermore, methods for editing the decay of an impulse response as well as the control of decorrelation were developed. The interaction dependent on the source parameter, named dynamic interaction, was studied in the main part of this chapter. User interfaces for the modification of source position and directivity as well as their automation have been studied. Prototype user interfaces for the interaction principles and visualization have been developed. Two different interfaces have been realized. A desktop user interface based on standard GUI elements using the MATLAB development environment was described. This implements the complete processing chain from simulation of array measurements using a mirror image source model and diffuse model to the interactive modification using the techniques developed in this thesis. An extended version using progressive interface technology was also implemented in a prototype system.

Chapter 7

Spatial Audio Reproduction

7.1 Introduction

This chapter discusses the reproduction of interactively modified sound fields. It corresponds to block **(3)** and **(4)** in Figure 7.1. Three different spatial sound reproduction systems are studied in detail.

1. Wave field synthesis
2. Stereophonic reproduction
3. Binaural reproduction

The focus of this work is the reproduction for a single user. These systems are chosen because they represent a broad range of different concepts. The aim of wave field synthesis is to reconstruct the sound field. It should be a solution without a sweet spot (i.e., not only a single point of correct reconstruction). This property is investigated in terms of room simulation for a single user. Stereophonic reproduction is a very common solution for loudspeaker reproduction with a sweet spot. More recently, different surround formats have been developed and some of these are already well-established (e.g., the 5.1 surround sound format). Since most of these systems can be modeled using a superposition of different stereophonic sets, only two-channel stereo reproduction is discussed in this context. The third system is the binaural reproduction, which is well-established in high quality simulation environments (Silzle et al., 2004).

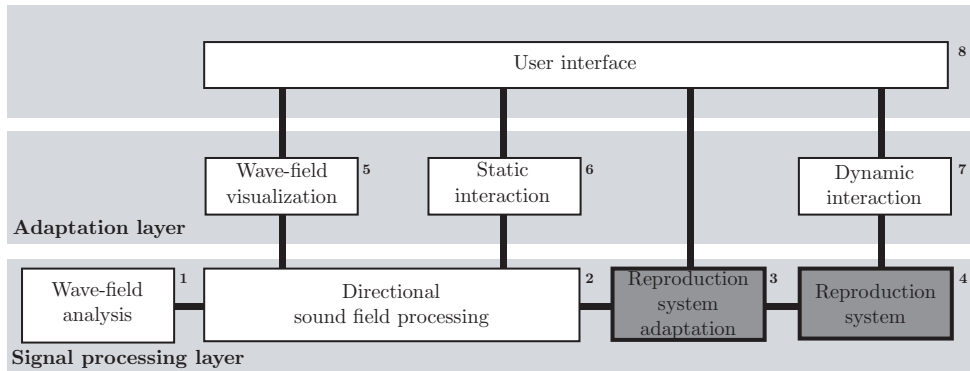


Figure 7.1: Basic block diagram of the system developed in this thesis. The parts discussed in this chapter are highlighted.

7.1.1 Outline of the Chapter

The chapter is structured into three main sections according to the three reproduction systems examined. Section 7.2 presents wave field synthesis (WFS) reproduction. After the derivation of the required synthesis operator, the implications of its limitations for auralization are analyzed briefly. An optimization of WFS for a single tracked user is proposed. The adaptation of the PWD to WFS reproduction is reviewed from the literature and discussed in view of the sound field analysis based on spherical array apertures. Section 7.3 analyses stereo reproduction. The concept of virtual microphones (VM) is introduced as a further sound design method. The adaptation of spherical array measurements to stereophonic reproduction is investigated. In Section 7.4 the binaural adaptation is studied. The theory of measurement and processing of head-related transfer functions (HRTFs) is given. Due to the limited resolution of publicly available databases, a set of high spatial resolution HRTFs of a dummy head was measured.

7.2 Wave Field Synthesis

Wave field synthesis is a method for the reproduction of a sound field within a certain area surrounded by loudspeakers. A listener moving in the reproduction area perceives virtual sound sources, which do not have to coincide with the real loudspeaker positions. Experiments to obtain such kind of reproduc-

tion can be found in the early publications of (Steinberg and Snow, 1934) and (Snow, 1953). They proposed the *acoustic curtain*. These early experiments can be interpreted as a way of implementing the Kirchhoff-Helmholtz integral. But the important step of synthetically generating the speaker signals was missing. The term *wave field synthesis* (often abbreviated as WFS) and the underlying principles have been formulated more recently by Berkhout (1988); Berkhout and de Vries (1989); Berkhout et al. (1993). The principles of wave field synthesis have been studied by various authors for different applications (e.g., acoustic enhancement by Vogel (1993), live sound enhancement by Start (1997), sound reproduction in general by (Verheijen, 1998), variable acoustics by Sonke (2000), video conferencing by de Bruijn (2004) and auralization by Hulsebos (2004)). In this section, the application of wave field synthesis for auralization in a larger reproduction area is considered. Limitations and perceptual effects of auralization using different reproduction schemes for direct sound, reflections and late reverberation are studied.

7.2.1 WFS Synthesis Operator

The theoretical basis for wave field synthesis is given by the Kirchhoff-Helmholtz integral (see Section 2.4). For practical implementation, a synthesis operator is required to calculate the loudspeaker (i.e., secondary source) signals for a desired configuration of virtual sources. Several methods exist to derive driving signals for the secondary sources to synthesize an acoustic field (Ahrens, 2010). The derivation of the synthesis operator requires several simplifications to arrive at a practical solution. These simplification have been studied in detail in (Vogel, 1993; Start, 1997; Verheijen, 1998; Hulsebos, 2004) among others. Most of the derivations simplify the Kirchhoff-Helmholtz integrals to a Rayleigh integral and lead to a solution for a line of secondary sources using the principle of stationary phase approximation twice. In this section, the derivation of the operator for point sources, focused sources and plane waves will be reviewed.

Point Source Synthesis Operator

The derivation given here for the synthesis of (virtual) primary point sources has been published by Start (1997). The corresponding geometry can be found in Figure 7.2. Here, only a short summary is given. The derivation steps lead

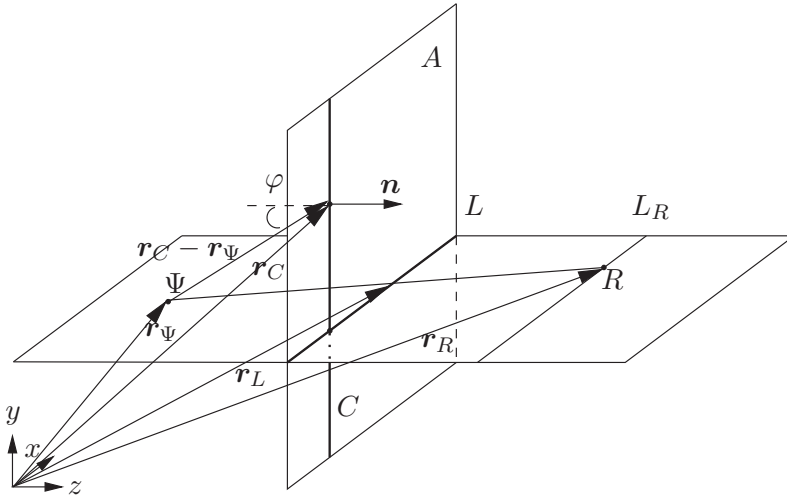


Figure 7.2: Geometry to derive the synthesis operator for a point source (Sladeczek, 2008).

to the synthesis operator $Q(\mathbf{r}_L, \omega)$ for one secondary source at the position \mathbf{r}_L on the line L . The pressure in the frequency domain $P(\mathbf{r}_R, \omega)$, in a receiver point R at position \mathbf{r}_R , can be calculated using the Rayleigh I integral

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_A \left[i\omega \varrho_0 V_n(\mathbf{r}_C, \omega) \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_C - \mathbf{r}_R|}}{|\mathbf{r}_C - \mathbf{r}_R|} \right] dA, \quad (7.1)$$

where \mathbf{r}_C denotes a point on the surface A , $V_n(\mathbf{r}_C, \omega)$ the normal velocity component in this point caused by a primary point source at location Ψ . This integral can also be written as an integration over all columns C of surface A along the x-axis,

$$P(\mathbf{r}_R, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P_C(\mathbf{r}_R, \omega) dx_C \quad (7.2)$$

where $P_C(\mathbf{r}_R, \omega)$ denotes the pressure in R generated by a single column C . $P_C(\mathbf{r}_R, \omega)$ can be calculated substituting the expression for the particle velocity of a monopole source (Start, 1997). Hence:

$$P_C(\mathbf{r}_R, \omega) = \int_{-\infty}^{\infty} S(\omega) i\omega \varrho_0 V_n \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_R - \mathbf{r}_C|}}{|\mathbf{r}_R - \mathbf{r}_C|} dy_C \quad (7.3)$$

with

$$i\omega\rho_0V_n = \cos\varphi \frac{1 + i\frac{\omega}{c}|\mathbf{r}_C - \mathbf{r}_\Psi|}{|\mathbf{r}_C - \mathbf{r}_\Psi|} \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_C - \mathbf{r}_\Psi|}}{|\mathbf{r}_C - \mathbf{r}_\Psi|}, \quad (7.4)$$

where φ defines the angle between the normal on the surface A and the vector $\mathbf{r}_C - \mathbf{r}_\Psi$ pointing from primary source (indicated by Ψ in Figure 7.2) to the secondary source position. $S(\omega)$ represents the primary source signal. In case the primary source Ψ and the receiver R are located in the same plane perpendicular to A , the stationary phase approximation can be applied. The basic idea is that the main contribution to the field in R is given by secondary sources along the line L . After applying the stationary phase approximation to the integral in Eq. 7.3, the surface A is reduced to the line L and $P_C(\mathbf{r}_R, \omega)$ becomes

$$P_C(\mathbf{r}_R, \omega) = H(\omega)S(\omega)g(\mathbf{r}_R, \mathbf{r}_L) \cos(\varphi) \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_L - \mathbf{r}_\Psi|}}{\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}} \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_R - \mathbf{r}_L|}}{|\mathbf{r}_R - \mathbf{r}_L|} \quad (7.5)$$

with

$$g(\mathbf{r}_R, \mathbf{r}_L) = \sqrt{\frac{|\mathbf{r}_R - \mathbf{r}_L|}{|\mathbf{r}_L - \mathbf{r}_\Psi| + |\mathbf{r}_R - \mathbf{r}_L|}}. \quad (7.6)$$

Using Eq. (7.5) the calculation of the loudspeaker driving functions $Q_R(\mathbf{r}_L, \mathbf{r}_R, \omega)$ for a specific receiver point R can be derived:

$$Q_R(\mathbf{r}_L, \mathbf{r}_R, \omega) = H(\omega)S(\omega)g(\mathbf{r}_R, \mathbf{r}_L) \cos(\varphi) \frac{e^{-ik|\mathbf{r}_L - \mathbf{r}_\Psi|}}{\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}} \quad (7.7)$$

$H(\omega)$ is an equalizing factor often described analytically by $H(\omega) = \sqrt{ik/2\pi}$. Beside this description, which is valid up to the aliasing frequency, source-dependent equalizing is still an issue of current research. In order to calculate the sound field in R , integration over the line L is required:

$$P(\mathbf{r}_R, \omega) = \int_{-\infty}^{\infty} Q_R(\mathbf{r}_L, \mathbf{r}_R, \omega) \frac{e^{-i\frac{\omega}{c}|\mathbf{r}_R - \mathbf{r}_L|}}{|\mathbf{r}_R - \mathbf{r}_L|} dx_L \quad (7.8)$$

Eq. (7.7) is only valid for a single listener or receiver position since the factor $g(\mathbf{r}_R, \mathbf{r}_L)$ depends on the position of the secondary source and the receiver. This equation can be transformed into an operator $Q(\mathbf{r}_L, \omega)$ for an exact reconstruction on a line of receiver points L_R , again using the stationary phase approximation. In this case, each secondary source defines a unique station-

ary phase point, which is found at the intersection point defined by the line connecting the (virtual) primary source position, receiver line L_R and the secondary source \mathbf{r}_L . In this example, the reference line L_R is taken parallel to the line L in Figure 7.2. The result of the stationary phase approximation is a different factor $g(\mathbf{r}_L)$:

$$g(\mathbf{r}_L) = \sqrt{\frac{|\mathbf{r}_R(\mathbf{r}_L) - \mathbf{r}_L|}{|\mathbf{r}_L - \mathbf{r}_\Psi| + |\mathbf{r}_R(\mathbf{r}_L) - \mathbf{r}_L|}}. \quad (7.9)$$

By substitution of $g(\mathbf{r}_R, \mathbf{r}_L)$ with $g(\mathbf{r}_L)$ in Eq. (7.7), a driving function for the receiver line L_R is defined.

$$Q(\mathbf{r}_L, \omega) = H(\omega)S(\omega)g(\mathbf{r}_L) \cos(\varphi) \frac{e^{-ik|\mathbf{r}_L - \mathbf{r}_\Psi|}}{\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}}. \quad (7.10)$$

This driving function enables the correct reconstruction of a wave field in terms of phase for the complete reconstruction area of the WFS system. The amplitude is only correct on the reference line L_R , which is not relevant for practical applications from a perceptual point of view. Depending on the secondary source layout used, the number of secondary sources required to synthesize a virtual primary source can vary. Besides a possible restriction of the reproduction area, a variance in the source loudness can occur. Methods for compensating for this effect as well as the required equalization are beyond the scope of this thesis. The reference line is not limited to a straight line. Instead, arbitrary curves may be used, as long as only one intersection with the reference line along each stationary phase line exists (Start, 1997; Sonke, 2000).

Focus Source Synthesis Operator

As shown in (Verheijen, 1998) and (Start, 1997), the synthesis operator can be modified to synthesize a focused source in front of the loudspeakers within the listening area using

$$Q_f(\mathbf{r}_L, \omega) = H(\omega)S(\omega)g_f(\mathbf{r}_L) \cos(\varphi) \frac{e^{i\frac{\omega}{c}|\mathbf{r}_L - \mathbf{r}_\Psi|}}{\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}} \quad (7.11)$$

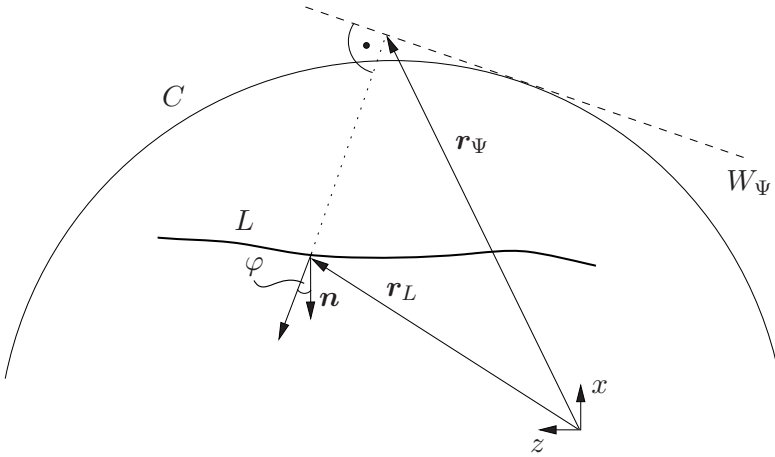


Figure 7.3: Geometry to derive the synthesis operator for a plane wave (Sladeczek, 2008).

with

$$g_f(\mathbf{r}_L) = \sqrt{\frac{|\mathbf{r}_R(\mathbf{r}_L) - \mathbf{r}_L|}{|\mathbf{r}_R(\mathbf{r}_L) - \mathbf{r}_L| - |\mathbf{r}_L - \mathbf{r}_\Psi|}}. \quad (7.12)$$

Plane Wave Synthesis Operator

To extend the synthesis operator to plane waves, a point source placed at an infinite distance to the listener is assumed. Taking the synthesis operator for point sources and letting $|\mathbf{r}_L - \mathbf{r}_\Psi|$ approach infinity, $g(\mathbf{r}_L)$ and the decay $1/\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}$ reach 0. This would result in zero as scale factor for secondary sources. The delay for the source represented in $e^{-ik|\mathbf{r}_L - \mathbf{r}_\Psi|}$ becomes infinite. For this reason, the plane wave operator is calculated in a different way. Let W_Ψ be the wavefront of a propagating plane wave with an angle φ to the normal \mathbf{n} on the secondary source array. The situation is shown in Figure 7.3. To calculate correct loudspeaker coefficients, a vector $\mathbf{r}_L - \mathbf{r}_\Psi$ orthogonal to the line W_Ψ to the secondary source is assumed. The synthesis operator for plane wavefronts can now be written as

$$Q_p(\mathbf{r}_L, \omega) = H(\omega)S(\omega) \cos \varphi e^{-i\frac{\omega}{c}|\mathbf{r}_L - \mathbf{r}_\Psi(\mathbf{r}_L)|}. \quad (7.13)$$

For the construction of the line W_Ψ , a circle C is used. The line W_Ψ is chosen as tangent on the circle C . This guarantees that all plane waves arrive at one

particular time in the center of the circle C . In a real system, C is chosen to be the center or reference point of the enclosed secondary source distribution. $H_p(\omega)$ is the equalization factor used for the plane wave reproduction. Furthermore, in all derivations of the synthesis operators, a loudness correction corresponding to the number of discrete secondary sources is required, which can also be integrated into $H_p(\omega)$. For further detailed derivations of the synthesis operator and analysis of the limitations due to sampling and truncation, the reader is referred to the literature (e.g., (Verheijen, 1998), (Vogel, 1993), (Start, 1997) and (Hulsebos, 2004)).

7.2.2 Single-user Optimized Wave Field Synthesis

The aim of WFS is to reconstruct the acoustic field within the entire listening area. Nevertheless, some effects cannot be perceived correctly in the entire listening area. The reasons are physical constraints as well as limitations of the practical implementation of a WFS system. In virtual reality systems (Springer et al., 2006), the user position has to be tracked in order to enable the correct image reproduction. In such a case the tracking data can also be used to optimize a WFS system. Three kinds of optimizations based on user tracking data have been studied and implemented:

1. Secondary source selection
2. Delay correction
3. Amplitude correction

These optimizations will be derived in the following section.

User-dependent Reference Line

In section 7.2.1 the WFS operator was derived. The form of the implementation presented requires a reference line for the calculation of the correct gain for the reproduction. In case of a linear secondary source distribution, the optimal solution of a reference line can be found by using a reference line parallel to the secondary sources. To optimize the wave field, the user distance from the line defined by the secondary source distribution can be used to modify the reference line. For WFS setups where the secondary sources are distributed around the listening area, an optimal solution for the receiver line is derived in

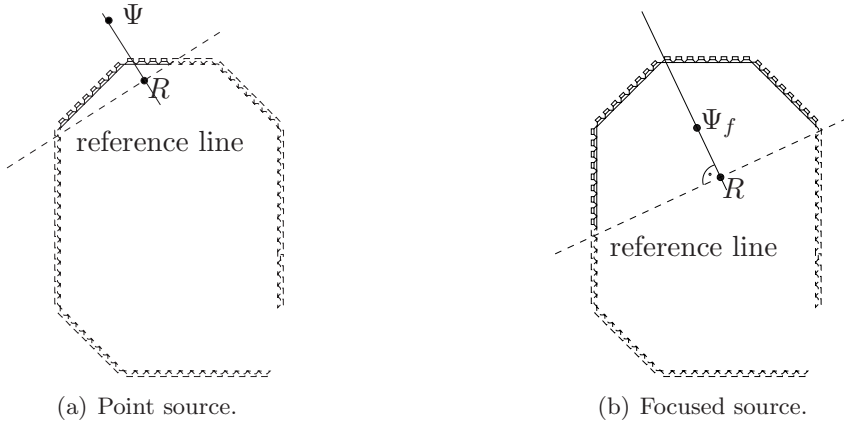


Figure 7.4: Selection of active speakers based on the reference line. Active speaker are marked in gray (Sładeczek, 2008).

(Verheijen, 1998). The receiver line is constructed as a line perpendicular to a line defined by the source and the center of the array (Spors, 2007). Instead of using the center of the secondary sources, the reference line can be constructed using a line perpendicular to a line defined by the actual source position and the position of the user. Figure 7.4 illustrates the construction of the reference line based on the position of the user.

Speaker Selection Optimization

The receiver line is used to select the secondary sources which are used to synthesize the virtual source. In order to get a large reproduction area of a source, it is advantageous if a large number of secondary sources contribute to the wave field synthesis process. The selection consists of two steps:

1. The receiver line is used to determine the potential candidates of secondary sources.
2. The angle φ between a vector \mathbf{r} pointing from the source to a speaker and the normal on the loudspeaker \mathbf{n} is used to consider whether the loudspeaker should be used for the synthesis. All speakers with $\varphi > 90^\circ$ are deactivated because this would invoke backpropagation toward the real source. are deactivated.

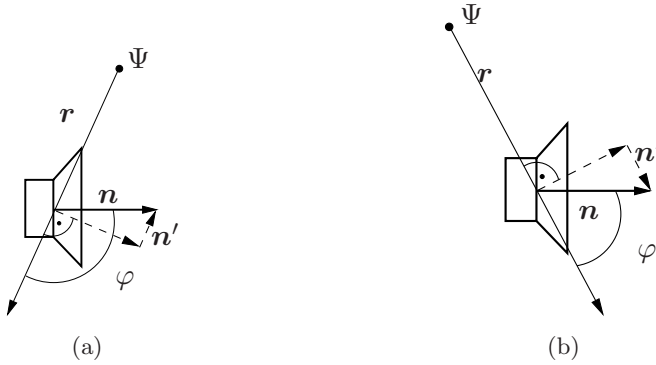


Figure 7.5: Selection of active speakers based on an angle criterion (Sladeczek, 2008).
 (a) Inactive speaker ($\varphi > 90^\circ$). (b) Active speaker ($\varphi < 90^\circ$).

Figure 7.4 presents an example for a point source Ψ using the first selection step. All speakers on the side of the virtual source are preselected (see (a)). For a focused source, the selection is done in an analog way as illustrated in Figure 7.4 (b). Then the second criterion as illustrated in Figure 7.5 is used to exclude secondary sources depending on their orientation. This method of speaker selection works for sources behind the array as well as for focused sources as illustrated. Another approach was given in (Spors, 2007), where an analytical source selection criterion is derived resulting in the same selection.

Time of Arrival Optimization

In specific applications, it is useful to minimize the distance-dependent delay of a source Ψ as much as possible. This is important if a WFS system is used for sound design purposes, instead of pure physical field reconstruction. In this scenario the sound designer wants to control the time of arrival of all sources independent of their spatial depth or distance. The WFS operator derived in Eq. (7.11) can be modified to keep the shape of the desired wave front and minimize the delay depending on the secondary source layout used. Figure 7.6(a) presents a typical WFS setup. A circle C with radius $|r_C|$ is constructed including all secondary sources. Now the timing of all virtual sources is optimized to arrive in the center of the array with the same delay, independent of their distance, while keeping the wave front curvature depending on their individual positions. The difference $|r_\Psi| - |r_C|$ can be used to reduce the delay in the

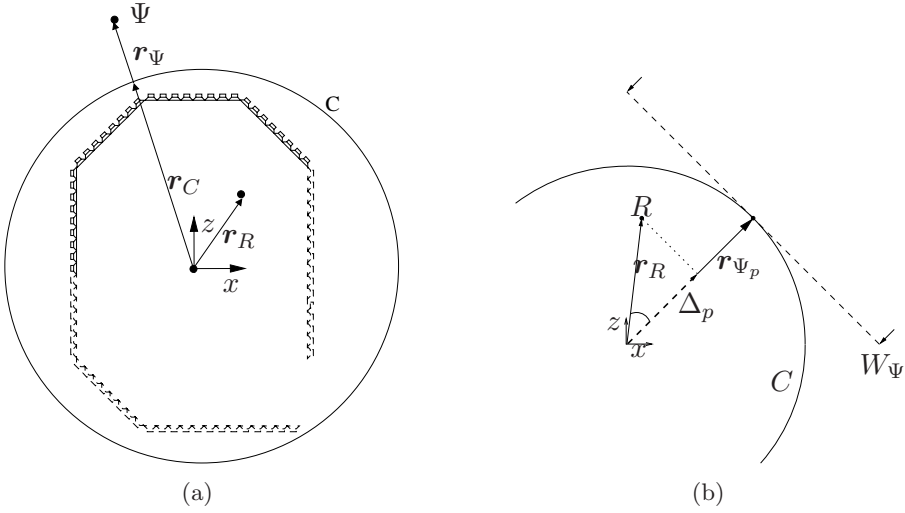


Figure 7.6: Geometry used to calculate the delay and level optimization for point sources (a) and plane waves (b) (Sladeczek, 2008).

synthesis operator by applying

$$d = |\mathbf{r}_\Psi| - |\mathbf{r}_C| \quad (7.14)$$

to the synthesis operator

$$Q(\mathbf{r}_L, \omega) = H(\omega)S(\omega)g(\mathbf{r}_L) \cos(\varphi) \frac{e^{-ik(|\mathbf{r}_L - \mathbf{r}_\Psi| - d)}}{\sqrt{|\mathbf{r}_L - \mathbf{r}_\Psi|}}. \quad (7.15)$$

In the next step the delay is kept constant for any user position \mathbf{r}_R using

$$d = |\mathbf{r}_\Psi| - |\mathbf{r}_C| - \underbrace{(|\mathbf{r}_\Psi| - |\mathbf{r}_\Psi - \mathbf{r}_R|)}_{\Delta} \quad (7.16)$$

$$d = |\mathbf{r}_\Psi - \mathbf{r}_R| - |\mathbf{r}_C|, \quad (7.17)$$

where Δ describes the distance difference of the listener with respect to the reference position in the center of the WFS system. It is also used to correct the level of a source. In the case of a plane wave, the geometry is depicted in Figure 7.6. A specific version of $\Delta = \Delta_p$ for plane waves can be calculated using the projection of the user vector \mathbf{r}_R on the vector \mathbf{r}_Ψ pointing to the

intersection point Ψ_p with the circle C

$$\Delta_p = \frac{\mathbf{r}_R \mathbf{r}_{\Psi_p}}{\|\mathbf{r}_{\Psi_p}\|} \mathbf{r}_{\Psi_p}. \quad (7.18)$$

As a result we get the user-optimized synthesis operator for a plane wave

$$Q_p(\mathbf{r}_L, \omega) = H_p(\omega) S(\omega) \cos \varphi e^{-ik(|\mathbf{r}_L - \mathbf{r}_{\Psi}(\mathbf{r}_L)| - \Delta_p)}. \quad (7.19)$$

Level Optimization

Using the synthesis operator derived in Section 7.2.1, the perceived level of a virtual source depends on the distance from the virtual source. In sound design applications as described in the previous section, it is useful to keep the level of a source constant for a listener moving inside the wave field. A desired level can be applied by a sound designer, adequate for the application (e.g., virtual reality environment). This is also useful for sources used for the reproduction of diffuse parts of a room. In order to synthesize such a source, the listener position \mathbf{r}_R is used to modify the synthesis operator. In the case of a virtual point source, the decay is $1/r$. To compensate for a different listener position, the factor $g_c(\mathbf{r}_{\Psi_p}, \mathbf{r}_R)$ is used with

$$g_c(\mathbf{r}_{\Psi}, \mathbf{r}_R) = \begin{cases} 1 + |\Delta| & \forall \Delta > 0 \\ \frac{1}{1+|\Delta|} & \forall \Delta < 0 \\ 1 & \forall \Delta = 0 \end{cases}. \quad (7.20)$$

In case of plane wave reproduction, the secondary source acts as a line source causing a $1/\sqrt{r}$ decay resulting in a correction factor given by

$$g_{cp}(\mathbf{r}_{\Psi_p}, \mathbf{r}_R) = \begin{cases} \sqrt{1 + |\Delta_p|} & \forall \Delta_p > 0 \\ \sqrt{\frac{1}{1+|\Delta_p|}} & \forall \Delta_p < 0 \\ 1 & \forall \Delta_p = 0 \end{cases}. \quad (7.21)$$

Directional Sources

The reproduction of directive sources through WFS was first proposed in the early work of (Verheijen, 1998). A direction-dependent factor $G(\varphi, \theta)$ is part

of the synthesis operator. If only a first-order directivity in a 2D system is required, $G(\varphi, \beta, \omega)$ is used with

$$G(\varphi, \beta, \omega) = \beta(\omega) - [1 - \beta(\omega)] \cos(\varphi), \quad (7.22)$$

where $\beta(\omega) \in [0, 1]$. A more general approach, which includes this elementary function, is the use of circular harmonics for the description of source directivity as described in (Corteel, 2007) and (Ahrens and Spors, 2007). If a data-based approach is used, filters for each direction of the source are stored in a database. These filters can be integrated into the synthesis operator. The limiting factor in all implementations is the aliasing frequency of the secondary source setup (Verheijen, 1998) (Vogel, 1993) (Start, 1997) (Hulsebos, 2004). The wave field can be controlled only up to this frequency. In the case of a point source emitting a broadband signal, the position is correctly perceived because sufficient energy is below the aliasing frequency. If the signal is high-pass filtered, a blur of the perceived localization is the consequence (Start, 1997). The simulated directivity of a source can only be explored for a limited frequency range. Proposals to optimize the high frequency processing have been made by (Wittek, 2002). He proposed the use of stereo reproduction above the aliasing frequency. This again leads to a sweet spot solution. However, if the field is optimized using tracking data, the stereo reproduction can also be optimized. In the proposed concept, $G(\varphi, \beta, \omega)$ is calculated in the driving function for each loudspeaker depending on the angle between the source and the corresponding speaker. If a directivity for a single user should be realized, $G(\alpha, \beta, \omega)$ can be calculated depending on the angle α between the normal of the source and the listener position. The situation is drawn in Figure 7.7 for a focused source Ψ_f and a listener R . This factor can be used to modify the source signal, independent of the secondary sources.

7.2.3 Reproduction Data for WFS Reproduction

The *high resolution data* generated during the wave field analysis process have to be adapted for reproduction using wave field synthesis. The result is a specific set of *reproduction data*. The WFS systems used in this research are all systems with loudspeaker arrays in the horizontal plane only. As a result the three-dimensional data of the spherical array analysis process has to be reduced to the horizontal plane. This can be done using the techniques described in Section 6.3.3 utilizing spatial windowing. Once the high resolution data have

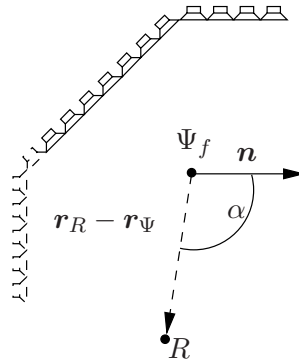


Figure 7.7: Geometry used to calculate the user dependent directivity factor $G(\alpha, \beta, \omega)$. Inactive speaker are drawn in dashed line (Sladeczek, 2008).

been reduced to the horizontal plane, one can consider different schemes for reproduction using a WFS system. For this purpose the three parts (discrete reflections, early part, late part) of the high resolution data are considered separately in terms of optimal reproduction properties. The processing scheme for the adaptation is depicted in Figure 7.8. The *discrete reflections* need to be perceived at a stable position for the whole reproduction area. Furthermore, in an optimal case they should be modified according to the source parameters (position, directivity) during reproduction. To ensure the availability of these properties, they should be kept separate and be reproduced as discrete point sources. In the case of static sources they can be added to the *early part*, which contains reflections and the first diffuse sound field elements. Like the discrete reflections, this part of the impulse response should be reproduced with a constant position for the entire audience. This can be achieved using point sources distributed around the listening area. The high resolution data have to be adapted to this resolution. The *late part* of the impulse responses contains the reverberation tail. (Sonke, 2000) evaluated the reproduction of a diffuse sound field perceptually. He showed that ten plane waves are sufficient to reproduce a rotation-invariant field for the listener in the center of the array. As shown in the following chapter this number is not sufficient if the user can move around in the listening area of the WFS system. Only if the field parameters are adapted to the user position, the reproduction is optimal. The number of plane waves must be adapted to the specific needs of a given scenario. The use of a plane wave instead of a point source for the late part of the impulses responses is optimal in terms of level distribution as shown in (Boone et al., 1999) and has been proposed by several authors (e.g., (Hulsebos,

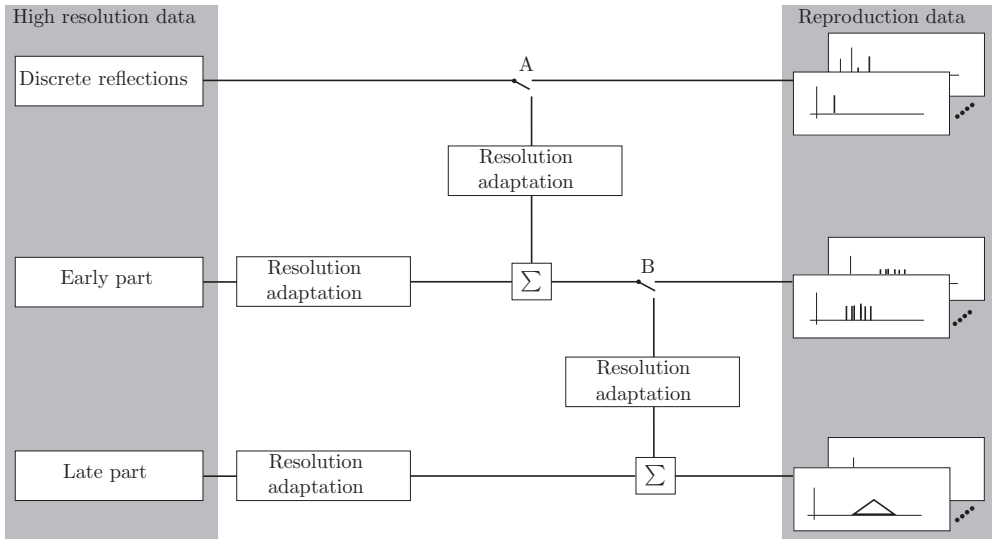


Figure 7.8: Adaptation of high resolution data to reproduction data for WFS reproduction

2004) (Sonke, 2000)).

In the case of limited computational power, the reproduction can be simplified further. The early data can be mixed to the late part, and only a limited number of plane waves are used with a single discrete source for the direct sound. These different possibilities are reflected in the reproduction scheme shown in Figure 7.8. In case of high computational resources only, a resolution adaptation is applied to all parts of the impulse responses. This results in a new set of impulse responses which can be used for the auralization. If the schema should be simplified as described above, the discrete reflections are added to the early part using switch *A*. If further simplification is required, the complete early part is added to the late part using switch *B*. This way a flexible adaptation is realized. This adaptation schema was also realized as part of the MATLAB user interface.

7.2.4 Dynamic Interaction

To auralize the *reproduction data*, specific audio processing is required. Figure 7.9 presents a block diagram of the dynamic interaction and reproduction process for WFS reproduction. The block diagram illustrates the signal flow

for a single audio source. If multiple sources in multiple rooms are to be auralized, the schema has to be applied for each source. In the following example, only one source and its signal paths are considered. All signal paths are equipped with an equalizer (EQ) to allow (optional) real-time modification of their characteristics. The audio stream is distributed to the appropriate number of signals according to the reproduction scheme. These are fed to FIR filters, which convolve the audio signals with the generated impulse response. The outputs of the separate signal paths are the input signals for the WFS rendering. While the direct sound and the discrete reflections can be realized as dynamic sources, the early reflections and late part are static source configurations. The source layout depends on the *reproduction data* as discussed in the previous section. The discrete reflections are positioned according to the pseudo-extrapolation process. To minimize the number of required WFS sources, the signals can be mixed to the signal of the early part using switch *A*. The summation can be done by a gain matrix using an intensity panning method (e.g., Vector based amplitude panning as described in (Pulkki, 1997)).

7.3 Stereophonic Reproduction

This section studies the adaptation of a plane wave decomposition to stereophonic systems. The adaptation is based on the simulation of a microphone setup. Since the impulse responses of such microphones are generated based on the plane wave decomposition of a room impulse response, they termed virtual microphones (VM). First, the general principles will be explained, then the influence of measurement errors from the plane wave decomposition process on the adaptation will be studied.

7.3.1 Stereophonic Microphone Setups

In the following section the simulation of virtual microphones by microphone array measurements and plane wave decomposition is studied. For this purpose the properties of ideal stereophonic microphone setups will be analyzed first. Since the invention of stereophonic sound reproduction in the early 1930's, a large number of microphone setups have been developed (Blumlein, 1931) (Alexander, 2000). Discussions on the different established techniques can

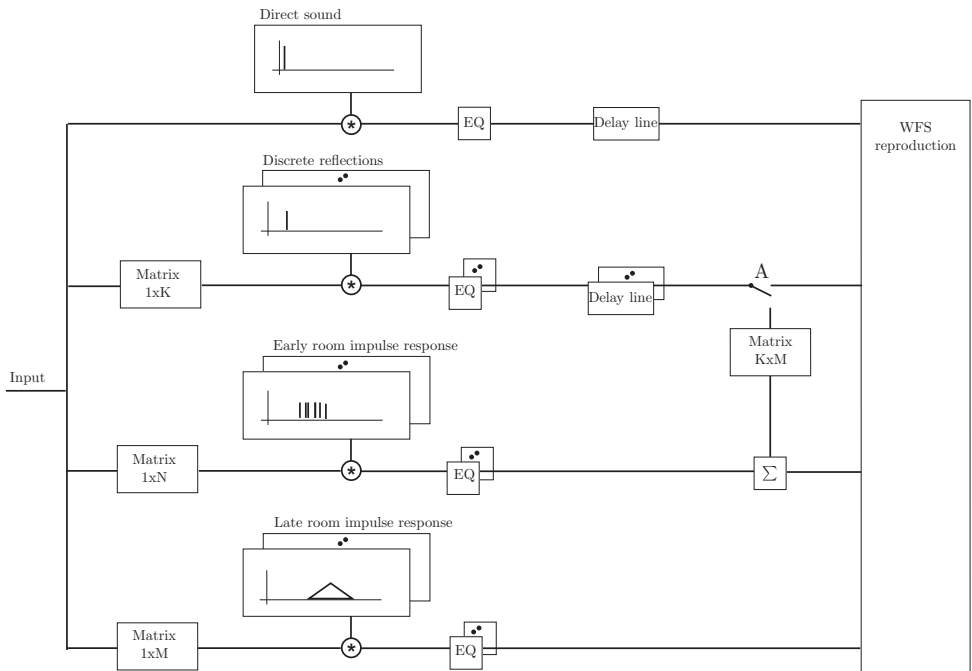


Figure 7.9: Block diagram of the processing required for the auralization and dynamic interaction for WFS reproduction.

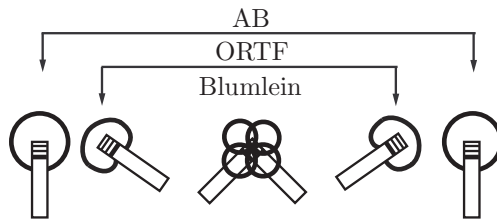


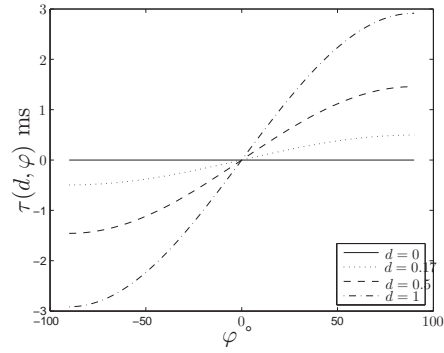
Figure 7.10: Geometrical layout of the stereo microphone setups investigated.

be found in (Bartlett, 1991), (Görne, 1996). For an overview of state-of-the-art surround recording techniques, the reader is referred to (Theile, 2001), (Williams and Du, 2000), and (Wuttke, 2008). In this thesis only two-channel stereophonic microphone setups are considered, but the outcome can easily be extended to common surround microphone techniques by a pairwise inspection of their reproduction channels. This investigation will be limited to three different setups: ORTF, AB and Blumlein stereo microphone setup (BL). Figure 7.10 presents a view of the three setups. The principles of stereo recording and reproduction are based on the creation of an inter-aural time difference (ITD) and/or an inter-aural level difference (ILD) for a listener positioned in an equilateral triangle with the two loudspeakers. This is done by including corresponding inter-channel differences in the signals of the speakers. The three microphone setups are chosen because they create pure time differences (AB), pure level differences (BL) and a combination of both (ORTF). To describe the characteristics of these setups, the following measures are used:

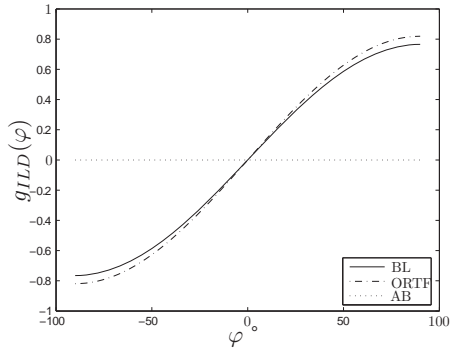
- Directivity pattern of a single microphone
- Function of inter-channel time differences and angle of incidence
- Function of inter-channel level differences and angle of incidence
- Spatio-temporal correlation coefficient and coherence function as defined in Section 2.7

The first three measures are important in relation to the reproduction of direct sound and discrete reflections. The last one describes the characteristic of diffuse sound reproduction.

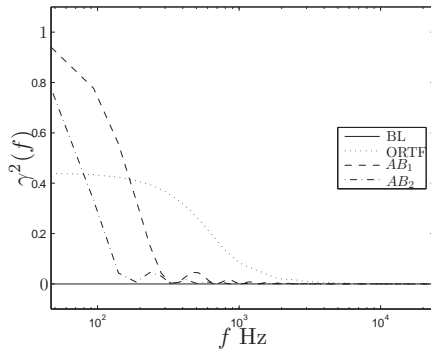
Inter-channel time difference This measure can be described analytically by simulating a single plane wave with a varying angle of incidence and plotting the parameter of the microphone setup. In case of an ideal two channel



(a) ITD



(b) ILD



(c) Coherence

Figure 7.11: (a) Inter-channel time difference for different distances d of a stereo omnidirectional microphone setup. (b) Difference of the level for the denoted microphone setups. (c) Analytical coherence function for different stereo microphone setups. AB_1 corresponds to an inter-microphone distance of $d = 0.5$ m and AB_2 using $d = 1$ m.

microphone setup in a plane, the inter-channel time difference $\tau(d, \varphi)$ can be calculated using

$$\tau(d, \varphi) = \frac{d \cos(\varphi)}{c}, \quad (7.23)$$

where d denotes the distance between the two microphones, c the speed of sound and φ the angle of incidence in the horizontal plane. An example for different microphone distances can be found in Figure 7.11(a).

Inter-channel level difference The inter-channel level difference $g_{ILD}(\varphi, \vartheta)$ depends on the directivity $g_{(\cdot)}(\varphi, \vartheta)$ of the two microphones. It can be calculated for the DOA with azimuth φ and elevation θ with

$$g_{ILD}(\varphi, \vartheta) = g_1(\varphi, \vartheta) - g_2(\varphi, \vartheta). \quad (7.24)$$

The ideal characteristics of ILD $g_{ILD}(\varphi, 0)$ and ITD $\tau(d, \varphi)$ of the chosen setups are given in figure 7.11(b).

Coherence measurements Another important property is the coherence characteristic of a specific stereophonic microphone setup. The correlation and coherence measurements are calculated based on the derivations in Appendix A.5, resulting in an analytical coherence function as given in Figure 7.11(c). In the practical realization of such setups, the characteristic is limited by physical constraint. For comparison Figure 3.20 on page 73 presents a polar plot of a real high quality microphone. It can be noticed that for higher frequencies, the desired polar pattern is not obtained.

7.3.2 Principles of Virtual Microphones

As mentioned before, the concept of virtual microphones is used in this thesis to generate a stereophonic reproduction based on array measurements. A virtual microphone is the result of an acoustic field in a given point captured with a given sensor including a specific directivity. Using a spherical array measurement, the acoustic field can be reconstructed in the volume of the sphere up to a given frequency. So the first step of generating a virtual microphone is an extrapolation of the field inside the sphere given by the array. The required principles have been studied in Chapter 5. Then a first-order directivity is applied to the extrapolated field components. This can be performed very simple

in the plane wave domain by a direction-dependent weighting of the plane wave components.

7.3.3 Reproduction Data for Virtual Room Microphones

Based on the *high resolution data* of spherical arrays, the *reproduction data* for stereophonic reproduction are generated. The process can be defined in general as presented in Figure 7.12. The *reproduction data* are then used in the dynamic interaction process, which is described in the next section. The impulse responses for the discrete reflections can be used directly in the dynamic interaction process. The early and late part have to be reduced to a single impulse response, corresponding to the desired microphone direction and directivity. For this purpose the wave field extrapolation based on the plane wave decomposition is used (see Section 5.2). First the field is extrapolated, then the weighting function, corresponding to the desired directivity is applied. The result is summed to achieve the impulse response of the virtual room microphone. This process requires a tremendous processing power, especially if the impulse responses are long and of an adequate spatial resolution to avoid spatial aliasing. For this reason an alternative approach for the late part of the RIR can be used. Based on the average of the measured late part, a single impulse response is generated. If a second impulse response is required, as in a stereo set-up, a diffuse part is generated based on the decorrelation with the desired coherence function as described in Section 2.7.2.

7.3.4 Dynamic Interaction

Based on the *reproduction data* as described in the previous section, the dynamic interaction processing leads to the stereophonic auralization of the desired room effect. A block diagram is presented in Figure 7.13. The input signal is distributed into four signal paths corresponding to the direct sound, discrete reflection, early part of the room impulse response, and late part of the room impulse response. The signals for the late part and the early part are convolved with two corresponding impulse responses stored in the *reproduction data*. These two signal paths are independent of the source parameter. Depending on the source position, the gains and delays of the direct sound and the discrete reflections are calculated. For the direct sound the required gains and delays can be calculated by simulating the desired virtual micro-

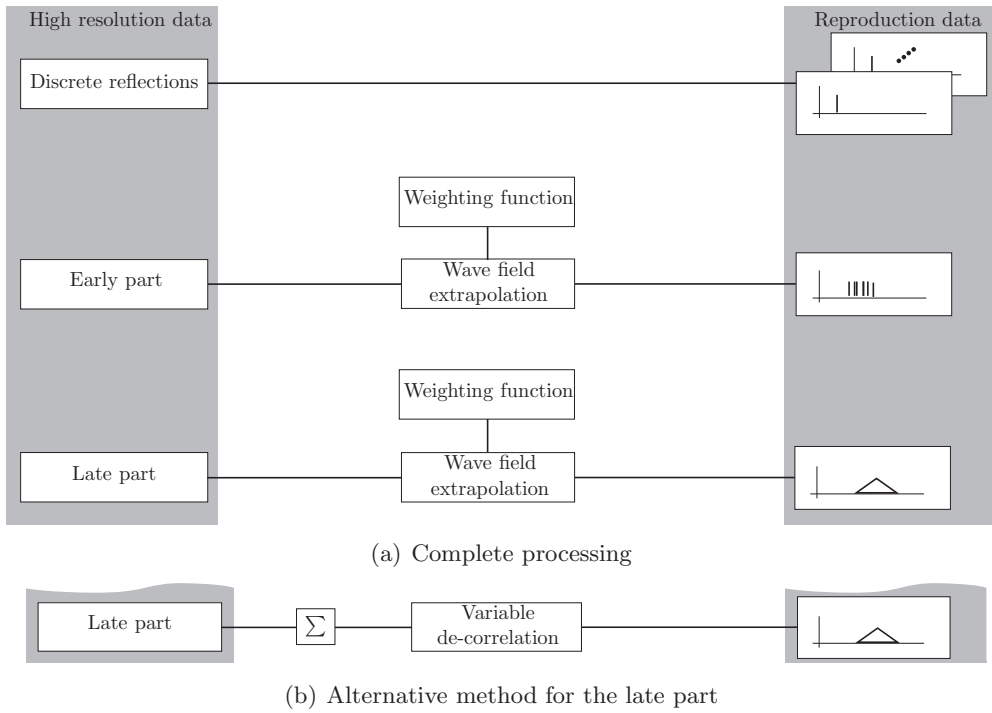


Figure 7.12: Adaptation of high resolution data to reproduction data for virtual microphone reproduction.

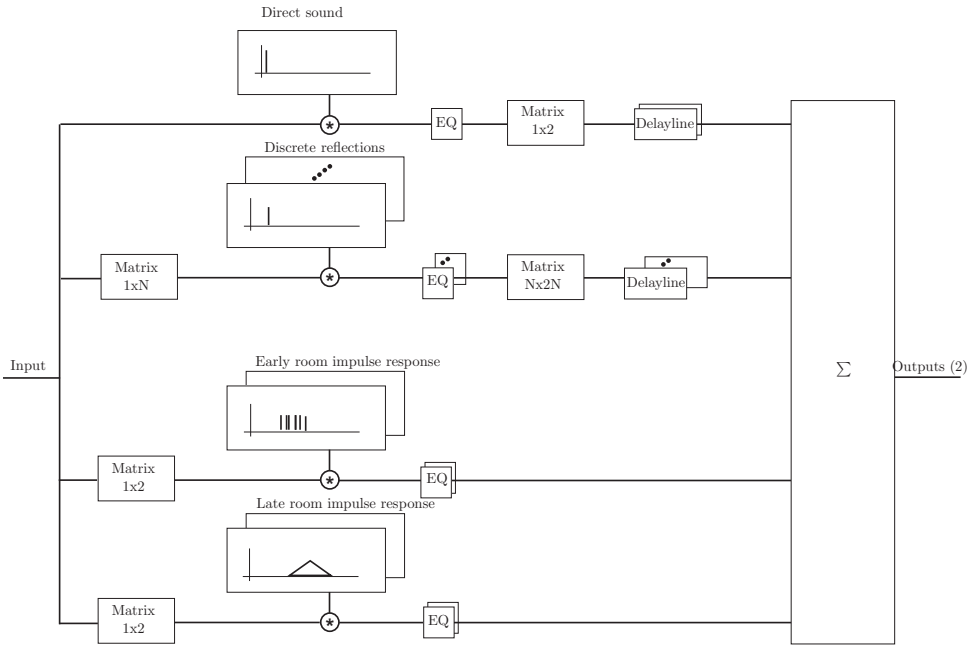


Figure 7.13: Block diagram of the processing required for the auralization and dynamic interaction for stereophonic reproduction.

phone setup. Such a virtual microphone-based panning tool was described by (Braasch, 2005). The direction-dependent inter-channel difference can be calculated using Eq. (7.23) and (7.23) on page 192. Depending on the distance of the source to the virtual microphone setup, an additional delay and gain reduction can be applied if a realistic behavior is desired. Often such parameters are separately controlled by the sound designer. The positions of the discrete reflections for a new source position are calculated using the pseudo-extrapolation model described in Section 6.4.4. A directivity of the source can also be included by applying an additional filter to the direct sound and including such a filter in the reflections. Alternatively, the equalizer (EQ) can be used to approximate the frequency characteristic of the reflections and direct sound (e.g., to simulate a distance-dependent air damping). From a real-time processing point of view, the use of an EQ based on an IIR-filter is more effective and easier to control. The stereo output signal is obtained by summing the different signal paths of the processing.

7.4 Binaural Reproduction

Besides the reconstruction of wave fields or the reproduction of dedicated binaural cues, the binaural reproduction using headphones leads to high quality spatial audio reproduction for a single user. In case of binaural synthesis, virtual sound sources are implemented with locations relative to the listener. This can be achieved by convolving an anechoic signal with a pair of head-related transfer function (HRTF) by reproducing the resulting signal using equalized headphones. An HRTF is defined as the sound pressure at the eardrum or at the ear channel entrance due to a sound source, divided by the sound pressure in the middle of the head with the head absent (Vorländer, 2008). It has been shown that the HRTFs are relatively independent of the source distance if this is larger than 1 m (Duda and Martens, 1998) but depend strongly on the direction of the source. Some authors have tried to model the spatio-temporal behavior of the ear using a beam-forming approach (Chen et al., 1992). In this thesis the sound field is analyzed using a plane wave decomposition, and the directional components are weighted with the corresponding HRTFs. The basic idea of using plane wave decomposition together with corresponding HRTFs is described in (DuraiSwami et al., 2005). Systems based on the use of a binaural simulated loudspeaker setup have been proposed by Falch et al. (2003).

HRTF Processing

The generation of a high quality HRTF data-set is not a trivial task. In this sub-section several issues related to HRTF processing and measurement are discussed. The results are used for the processing of the HRTFs which were used for the listening experiment (described in Chapter 8) and for adaptation of a plane wave decomposition to binaural reproduction.

HRTF-preprocessing The measurement of head-related impulse response (HRIR) is influenced by the non-ideal characteristics of the measurement environment and the transfer function of the loudspeaker and microphone used. Such distortions can be reduced by appropriate corrections. An overview of existing techniques can be found in (Larcher et al., 1998). In the current work diffuse-field equalized impulse responses are used. The diffuse-field HRTFs magnitude spectrum can be estimated by measurements in a reverberant chamber, average frequency response measurements in a free field (anechoic room)

or transient excitation in a diffuse field (typical room). Because only HRTFs measured in an anechoic chamber are available, in this research the second method is used. The pressure reaching the listener's eardrum is given in the frequency domain as $P(\omega)$ and the source signal as $S(\omega)$. If measured in a diffuse sound field, $P(\omega)$ can be expressed as

$$P(\omega) = S(\omega) \cdot H_{diff}(\omega). \quad (7.25)$$

The magnitude of the diffuse-field HRTF can be approximated with

$$H_{diff}(\omega) = \frac{P(\omega)}{S(\omega)} = \sqrt{\sum_n |H_n(\omega)|^2}. \quad (7.26)$$

The filter $H_{diff}(\omega)$ is used to equalize the HRTFs measured in an anechoic room. This method relies on a large number of measurements. These cannot be reduced without limiting the robustness of the estimation (Larcher et al., 1998). Furthermore, for the subsequent processing it is worthwhile to separate the HRTFs into their components. A measured HRTF in the frequency domain $H(\omega)$ can be decomposed into a minimum phase component $H_{min}(\omega)$, an all-pass component $H_{all}(\omega)$ and a linear phase component $H_{lin}(\omega)$:

$$H(\omega) = H_{min}(\omega) \cdot H_{all}(\omega) \cdot H_{lin}(\omega) \quad (7.27)$$

The minimal phase component has the smallest possible phase angle and a magnitude exactly as the original transfer function. The linear phase component has a unity magnitude and corresponds to a pure delay. The all-pass component has a magnitude of unity and carries the remaining phase of the original measurements (Minnaar et al., 1999). The minimum phase component can be obtained from the original measurement $H(\omega)$ using the cepstral method (Huopaniemi and III, 1999). Following (Oppenheim et al., 1999), a minimum-phase reconstruction of any system function can be carried out based on the property that the real and imaginary parts of $H(\omega)$ are related by the Hilbert transform. A minimum-phase version $|H_{min}(\omega)|$ of the magnitude response $|H(\omega)|$ can be found by windowing the real cepstrum. The excess phase component can be acquired using

$$H_{ex}(\omega) = \frac{H(\omega)}{H_{min}(\omega)} = H_{all}(\omega) \cdot H_{lin}(\omega). \quad (7.28)$$

The excess phase component carries the inter-aural time difference (ITD). It is desirable to extract the ITD from the excess phase component. Several approaches have been described in the literature. (Jot et al., 1995) proposed a method for determining the ITD by fitting a linear curve to the phase of the excess-phase component between 1 kHz and 5 kHz. A similar method was used by (Huopaniemi and III, 1999) in the frequency range of 500 Hz to 2 kHz. (Kistler and Wightman, 1992) calculate the ITD corresponding to the maximum in the cross-correlation function between measured HRTF pairs. Another method described by (Sandvad and Hammershoi, 1994) is to search for the first point in time the impulse response exceeds 5% of its maximum value for the left and right HRTF and subtract the two values to obtain an estimated ITD. Detailed description of the different methods is given in (Minnaar et al., 2000).

HRTF-interpolation In the field of binaural reproduction, the interpolation of HRTFs has been extensively investigated. The reason is that the desired resolution of HRTFs as required for interactive simulations is difficult to achieve because such measurements are very time-consuming (e.g., for a resolution of 5° on a sphere, 2450 measurements are required). Minnaar et al. (2005) investigated the required minimal directional resolution of HRTFs. Based on a criterion derived from listening experiments, the number of 11975 measured HRTFs for a resolution of two degrees was reduced to 1130. Publicly available HRTF databases have amounts from 187 (Paris, 2008) to 1250 (Algazi et al., 2001) directions. These do not correspond with the resolution required in the current work. Furthermore, the distribution of measured direction is often limited by practical constraints.

HRTF-measurements In this work the influence of different methods for HRTF-processing had to be minimized. For this reason a high resolution HRTF measurement on a dummy head was performed in the anechoic room at Fraunhofer IDMT, Ilmenau. The available dummy head was a Cortex MKI including Microtech Gefell microphones. The measurement was performed using the measurement system Monkey Forest including an automated positioning device for the dummy head. A Klein & Hummel O110 loudspeaker was used at a distance of 4 m. Figure 7.14 presents the measurement setup with (the back of) the loudspeaker on the right and the dummy head on the positioning devices on the left side. The measurement was performed at a 2° resolution for azimuth and elevation, resulting in 16200 positions. For the measurement signal a logarithmic sweep of 1.36 s length was applied. In the post-processing the



Figure 7.14: Setup used for the HRTF-measurements.

measured impulse responses were diffuse-field equalized and reduced to 1024 samples. The diffuse-field filter was estimated using the methods described above using a minimum phase filter. This data is used in the subsequent processing and the listening experiment in the following chapter.

7.4.1 Reproduction Data for Binaural Reproduction

Based on the *high resolution data* and an HRTF-set, the *reproduction data* for binaural auralization could be generated. A block diagram of the generation is given in Figure 7.15. The adaptation is given for each of the three parts (discrete reflections, early room impulse response and late room impulse response). For the discrete reflections no further processing is required at this stage. The early part of the room impulse response plane wave decomposition $\bar{P}(\phi, \theta, \omega)$ is adapted to the required resolution M of the HRTF database using an adequate windowing method. Afterwards each plane wave component is convolved with the corresponding HRTF and summed for each ear:

$$P(\omega) = \sum_{i=0}^I \sum_{n=0}^N H_{HRTF}(\phi_i, \theta_n, \omega) w_s(\phi_i, \theta_n, \omega), \quad (7.29)$$

where $w_s(\phi_i, \theta_n, \omega)$ denote the discrete direction weights, $H_{HRTF}(\phi_i, \theta_n, \omega)$ the HRTF for the direction and $P(\omega)$ the desired transfer function. This process has to be performed for both ears and for all components of the *high resolution data*. There are several variables in the process of generating reproduction data. The resolution of the plane wave decomposition of the early part of the room impulse response and the resolution of the late part have to be adapted to the resolution of the HRTF database. For the early room impulse response part, a high resolution is necessary because of the required accuracy of positions of early reflections. The late part of the room impulse response can be adapted to a lower resolution. Mommertz (1996) has shown that a minimal resolution of 30 directions is required for the reproduction of the early part of the room impulse response, while for the late part the resolution can be reduced to 6 directions. To find the optimum for the system described in this work, further listening experiments are required to verify these results for the conditions found here.

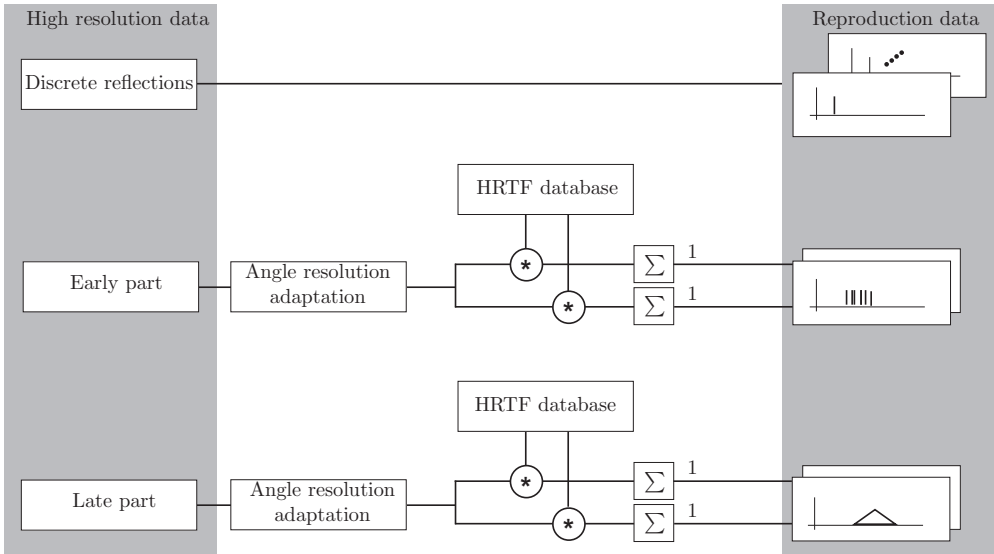


Figure 7.15: Adaptation of *high resolution data* to *reproduction data* for binaural reproduction.

7.4.2 Dynamic Interaction

Based on the *reproduction data*, the dynamic interaction can be implemented. Figure 7.16 presents the signal processing required for the dynamic interaction process for binaural reproduction. The incoming audio signal is distributed to four signal paths. The paths for the early and late part of the room impulse response only have to be convolved with the HRIR generated during the adaptation of the high resolution data to the reproduction data. To enable the dynamic interaction, the direct signal and the discrete reflections have to be convolved with HRIRs at this stage. The HRTFs are chosen from a database according to the required direction for the reflections and the direct sound. To simulate the distance, an appropriate delay and air damping can be included using the delay line and the equalizer (EQ). The positions of the discrete reflections are calculated according to the pseudo-extrapolation derived in Section 6.4.4. The result is summed into two output signals for the two ears.

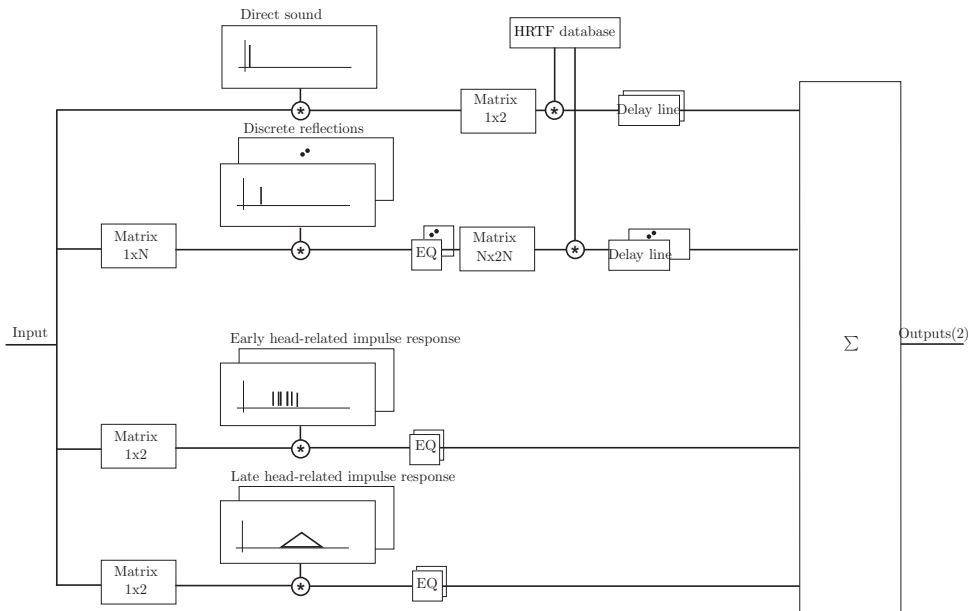


Figure 7.16: Block diagram of the processing required for the auralization and dynamic interaction for binaural audio systems.

7.4.3 Binaural Reproduction Using Head-tracking

The dynamic interaction can also be used in an environment with a head-tracked user. It has been shown that the externalization and perceived quality of a binaural reproduction is significantly better using head tracking (e.g., (Begault et al., 2001) (Minnaar et al., 2001)). In this case a higher number of binaural room impulse response sets is required for the late and early part of the room impulse response. These can be generated easily by altering the HRTF set used in the generation of the reproduction data, corresponding to the desired viewing direction of the listener. During the auralization the dataset used for the convolution is altered according to the tracked head orientation of the listener. It is possible to be extended this to all different rotation axes of the listener.

7.5 Summary

This section studied the adaptation and reproduction of spatial high resolution data, measured using a spherical array, to different reproduction systems. For the WFS adaptation and reproduction, a user-optimized system was developed, successfully implemented, and verified. The system optimized the reproduction of focused sources and diffuse fields for a single listener. A processing scheme and a data adaptation scheme for high resolution room impulse responses was developed. The second reproduction system examined was the stereophonic reproduction. An adaptation scheme and a processing scheme for *high resolution data* and the corresponding *reproduction data* was developed. The third spatial reproduction system studied was the binaural reproduction. The adaptation and reproduction of the measurements was examined as well. A high resolution set of HRTFs were generated.

Chapter 8

Perceptual Evaluation

8.1 Introduction

The first part of this chapter (Section 8.2) focuses on the perception of early reflection according to their relevance for the correct impression of source position and source distance. Two pilot listening experiments have been conducted to support the assumption that the modification of only a few discrete reflections is sufficient to simulate the main source parameter: position, directivity and distance of a source. Pellegrini (2001) studied the use of a limited number of artificial reflections to stimulate the perception of distance and room size. In the context of this thesis the source position and directivity were studied. The aim was to verify if the modification of a limited number of strong first-order reflections is sufficient to generate a high quality auralization. In the first experiment the source directivity was studied (Section 8.2.1); in the second experiment the source position was studied (Section 8.2.4). Both experiments use the same simulation environment as described in the next section. The second part of the chapter discusses the perception of sound fields generated by adaptation of spherical array measurements to different reproduction systems as examined in the previous chapter. Three different spatial sound reproduction systems are studied in detail: wave field synthesis, stereophonic reproduction and binaural reproduction. In Section 8.3 a prototype system of a single-user optimized WFS system is evaluated. The reproduction quality of single-user optimized focused sources and single-user optimized diffuse field reproduction is studied. Section 8.4 analyzes stereo reproduction. A quality grading experiment was made to compare simulated stereo microphone setups with the virtual microphones (VM) generated for simulated spherical arrays with and without measurement errors. In Section 8.5 the binaural adaptation is studied. A spherical array measurement with and without measurement

errors was simulated. Finally several methods to adapt the simulations to a binaural reproduction were compared with an ideal binaural simulation in a listening experiment.

8.2 Influence of Correct Reproduction of Early Reflections

General consideration The two experiments conducted are both using a binaural simulation environment. To minimize the influence of the Head-Related Transfer Functions (HRTFs), individualized HRTFs are used. For this purpose, a pre-test was performed to select the best fitting HRTF for each subject. The HRTFs are taken from the CIPIC database (Algazi et al., 2001) and were diffuse-field equalized. The subjects have the possibility to listen to 45 different HRTFs convolved with a test signal. The test signal is 30 Hz amplitude modulated white noise. Three cycles of these bursts are convolved with an HRIR corresponding to a horizontal direction before switching to the next direction. A virtual sound source is moved two times around the listeners head. Subjects are asked to choose the HRTF set with no elevation and a smooth frequency response during the movement, as well as with the best externalization especially for the frontal direction.

Room model To generate the examples for the listening test, a mirror image source model (MISM) was used. The aim was not to generate a very realistic environment, but to try to analyze the worst case. It is expected that in an image source model without any diffraction and diffuse tail, the differences between the different reflection patterns are most audible. In other words, if in this situation a given order of correct reflections is sufficient than this order may also be sufficient for real environments. The implementation is analogous to the description by Allen and Berkley (1979) and Borish (1984). To extend the image source model to directive sources, a first-order directivity was implemented. The directivity can be described by:

$$g(\varphi, \vartheta, \beta) = \beta + (1 - \beta)\cos(\varphi)\cos(\vartheta) \quad (8.1)$$

with

$$\beta \in [0, 1]$$

The $g(\varphi, \vartheta, \beta)$ is used to express the frequency-independent directivity depending on the azimuth angle φ and the elevation ϑ , using β to vary the directivity between figure-of-eight ($\beta = 0$) and omni-directional characteristic ($\beta = 1$). As an extension a *first-order frontal* characteristic was used. In this case $g(\varphi, \vartheta, \beta)$ is set to zero for all $\pi > \varphi > 2\pi$. A binaural impulse response $P(\omega)$ from an image source model is calculated for each ear using:

$$P(\omega) = \sum_{n=1}^N H_{HRTF}(\omega, \varphi_n, \vartheta_n) H_{ref}(\omega, \varphi_n, \vartheta_n) \alpha_{air}(\omega, r_n) g(\varphi_n, \vartheta_n, \beta), \quad (8.2)$$

where N reflections at the positions $(\varphi_n, \vartheta_n, r_n)$ are summed. With a frequency response $H_{ref}(\omega, \varphi_n, \vartheta_n)$, an air damping factor $\alpha_{air}(\omega, r_n)$ given by Eq. (A.25), a direction factor $g(\varphi_n, \vartheta_n, \beta)$ and the HRTF corresponding to the direction $H_{HRTF}(\omega, \varphi_n, \vartheta_n)$, the binaural impulse response $P(\omega)$ is calculated.

8.2.1 Perception of Source Direction

The orientation of a directional sound source in a anechoic environment or in a system reproducing direct sound only is difficult to perceive for a listener. If the listener can move, one can explore the different radiation of a source depending on its listening position. In case of a static listener one will only perceive a difference in the source timbre itself without affecting the impression of different directivity or the orientation of a source. To make the directivity of a source perceivable even for a static listener in a reproduction system, room simulation is required. In the experiment the importance of correct auralization of reflections generated by the mirror image source model (MISM) described above is investigated.

8.2.2 Listening Experiment

For the listening experiment all items are generated with specific selected HRTFs for each subject as described above. During a training phase, listeners have the possibility to listen to a sound source with different orientations using the same directivity as in the later test. The source is rotated in azimuth with steps of 45 degrees during this training phase. The hypothesis of this listening experiment was:

To achieve a good quality in perception of source directivity and orientation, it

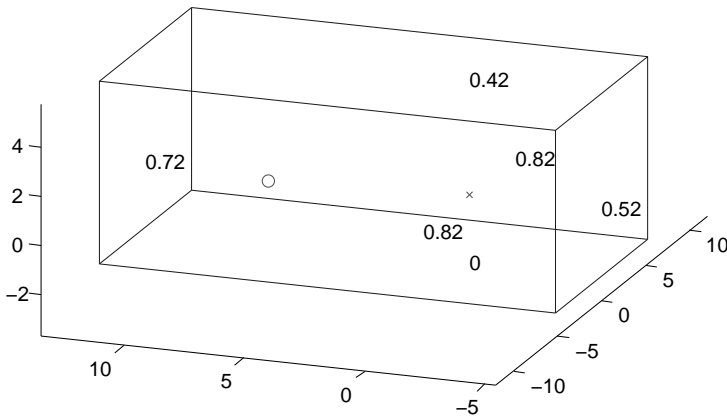


Figure 8.1: Simulated room with reflection coefficients for each wall. The source is marked with a o , the receiver is marked with a x . All dimensions are given in m.

is sufficient to generate only the first-order reflections in a correct way. Reflections of higher order can be calculated assuming an omnidirectional source.

To proof this hypothesis, a MUSHRA quality grading experiment was performed. The environment for the test was a studio room. The signal chain with a PC includes a high-quality audio interface and STAX Lambda Pro electrostatic headphones. The MISM as described in the previous section was used to generate the binaural impulse response of a rectangular room, which is shown in Figure 8.1. The reflection coefficients are frequency-independent and presented at the corresponding wall in the figure. Two different source signals with different directivities are used while the room conditions are kept constant. As a source signal, one sentence of male speech is used (speech). As a second source signal, a mixed sequence of drums and an accompanying electric bass is used (drums). For each signal four source orientations (A, B, C, D) are tested (0° , 90° , 135° , 292°). 90° corresponds to the direction of the right wall from the listener's point of view. Beside these orientations, an anchor with the same source directivity pointing in the opposite direction with respect to the reference was always included. For the speech signal a *first order frontal* directivity was used. For the drums signal the simulation was divided into two frequency ranges. Up to 400 Hz an omnidirectional directivity is calculated and above 400 Hz the same first-order frontal directivity as for the speech sample was applied. For each of the four directions, five different items are rated (hid-

	Direct	1st Order	2nd Order	3rd Order
Reference	+	+	+	+
Item 1	+	+	+	o
Item 2	+	+	o	o
Item 3	+	o	o	o

Table 8.1: Overview of reflection rendering. + = Reflections rendered correct, o = Reflection rendered based on omnidirectional source.

den reference, anchor, item 1, item 2 and item 3). Each item was only correctly simulated up to a specific order of reflections. Table 8.1 presents an overview. The energy for the reflections is kept constant for the directional source and the omni-directional simulation. The task for the subject was to grade the quality of different items of one source direction on a five point scale. The grading was performed in direct comparison to an open reference. The subjects are free to listen to all other examples, which are presented simultaneously without a time limit. Each grading was performed twice by each subject, while the order of the items per signal was randomized.

8.2.3 Results

The subjects were trained listeners, but unexperienced with binaural simulations. For this reason the first gradings of the subjects are omitted. The results for the different signals are analyzed separately. Eight subjects, two female and six male in the age of 23 to 32, participated in this listening experiment. Normal hearing was reported by all subjects.

Speech example Figure 8.2(a) shows the results for the speech example. In the overall rating significant results are obtained by the hidden reference, item 3 and the anchor. Item 2 and item 3 show significant difference in comparison to the hidden reference, but are both in the excellent range. The difference in the rating between item 1 and item 2 is not significant. From this experiment it can be concluded that for the correct perception of source orientation under the given conditions, it is sufficient to model only the first-order reflections correctly and render all higher-order reflections under the assumption of a omni-directional source. Comparing the different source orientations, it is obvious that subjects could not distinguish the difference between the exam-

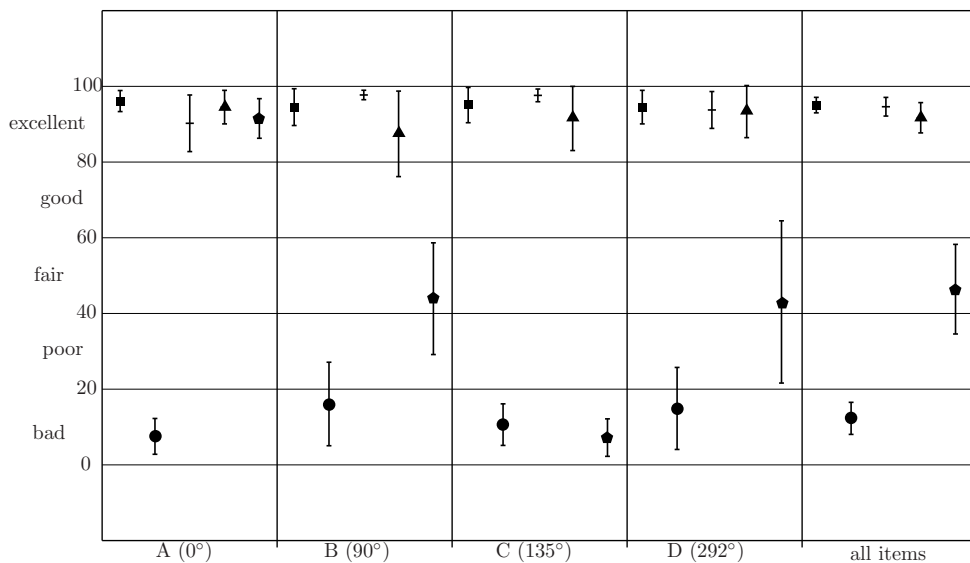
ples for orientation A. Only if the source is rotated, a difference is perceived and a grading becomes possible.

Drums example Figure 8.2(b) shows the results for the drums example. The overall tendency of the results is similar to the speech signal. Remarkable are the results for direction A and B. In this case there is a tendency to grade item 1 and item 2 differently. This is due to the fact that the drums example contains more transients, which makes it possible to identify flutter echoes easily in the room, especially for the given simple image source model. If second- and higher-order reflections are not modeled properly, these flutter echoes are not generated and a clear difference compared to the reference can be perceived.

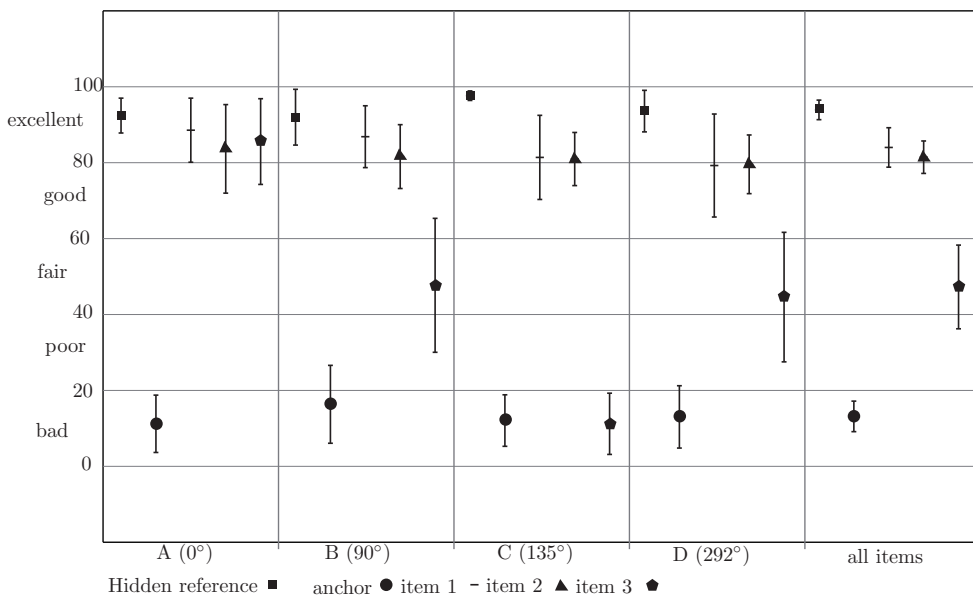
Conclusions For each signal the difference in the grading between different items depends on the source orientation. The anchor is nearly always identified as well as the hidden reference. To get an excellent or good grading for the source directivity, the correct simulation of the first-order and second-order reflections is sufficient. This holds for a speech signal and a broadband directivity simulation as well as for a drums example with a directivity simulation above 400 Hz. Correct simulation of only the direct sound results in a bad grading of the examples. It can be concluded that for the simulation of source directivity, the first-order and second-order reflections are most important.

8.2.4 Perception of Source Position

In this experiment the influence of the correct calculation of early reflections according to different source positions is investigated. In the previous experiment it was shown that only this first-order reflections have to be simulated according to the source orientation to get a good result in a quality grading. The grading was performed between the original sound field and a sound field in which only first-order reflections were rendered according to the source orientation and directivity. All higher-order reflections are rendered under assumption of an omnidirectional source. Another important factor, which influences the first-order reflections, is the position of the source. Helleman (Helleman, 2003) has performed experiments on the sensitivity of the human auditory system to spatial variations in single early reflections. She draws the conclusion that a source can be moved in a certain area without adapting the positions of the



(a) Speech example



(b) Drums example

Figure 8.2: Average results and 95% confidence intervals for the source orientation experiment.

early reflections. The area is defined by the threshold for spatial variations in the reflections. To take these investigations one step closer to a real situation, a listening experiment was designed. In this experiment the importance of correct auralization of different orders of reflection generated by a MISM is investigated for different source positions.

Listening experiment To generate the examples for the listening test, the same mirror image source model as in the previous experiment was used. Only omni-directional sources were simulated. To minimize the influence of the HRTF, individual selected HRTFs were used. The procedure was the same as in the first experiment. In fact, most of the subjects of the first experiment also performed this experiment. Before the subject started the grading experiment, a training was performed. During the training phase, listeners had the possibility to listen to different source positions of the later test. For each signal two variances were presented. The first example was calculated like the reference in the later test, while the second example was calculated like item 3 of the main test. The hypothesis of this listening test was:

To achieve a good quality in perception of source position in terms of room impression, it is sufficient to generate only the first-order reflections in a correct way. Reflections of higher order can be generated under the assumption of a static omni-directional source position in the center of the moving area.

To proof this hypothesis, a MUSHRA (ITU, 2003) quality grading experiment was performed. The test was carried out in a studio environment using a signal chain with a PC equipped with a high-quality audio interface and STAX Lambda pro headphones. The MISM was used to generate a binaural impulse response of a rectangular room, which is shown in Figure 8.3. Two different dry source signals were tested. The first source signal was one sentence of male speech (speech), and the second source was a short sequence of a drums with accompanying bass (drums). For each signal four positions (B, C, D, E) were tested. Position A is the initial position of the source (see Table 8.2). Beside this position an anchor with a different room size was also included. For each of the four positions five different items are rated (hidden reference, anchor, item 1, item 2 and item 3). Each item was simulated correctly up to a given order of reflections as given in Table 8.3. Each grading was done once by each subject, while the order of the items per signal was randomized. Two female and six male subjects in the age of 23 to 32, participated in this listening experiment. Normal hearing was reported by all subjects.

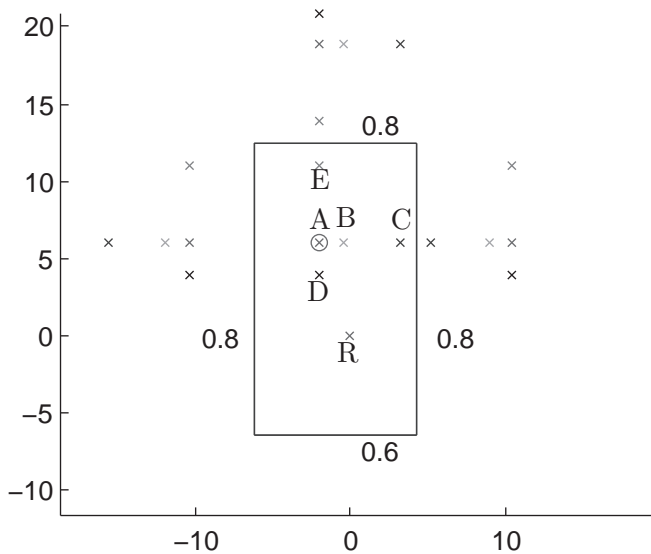


Figure 8.3: Rooms with reflections coefficients, source positions (A to E) according to Table 8.2, receiver position (R) and first order reflections used for the listening experiment. The ceiling had a reflection coefficient of 0.5 and the floor of 0. Only first order reflections for side walls and back wall are shown.

Position	φ	ϑ	r
A	18.43°	0°	6.32 m
B	4.76°	0°	6.00 m
C	-28.07°	0°	6.80 m
D	26.56°	0°	4.47 m
E	10.3°	0°	11.18 m

Table 8.2: Positions of the sources (A to E) with respect to the receiver position.

	Direct	1st Order	2nd Order	3rd Order and higher
Reference	+	+	+	+
Item 1	+	+	+	o
Item 2	+	+	o	o
Item 3	+	o	o	o

Table 8.3: Overview of reflection rendering. + = Reflections rendered correct, o = Reflection rendered based on position A

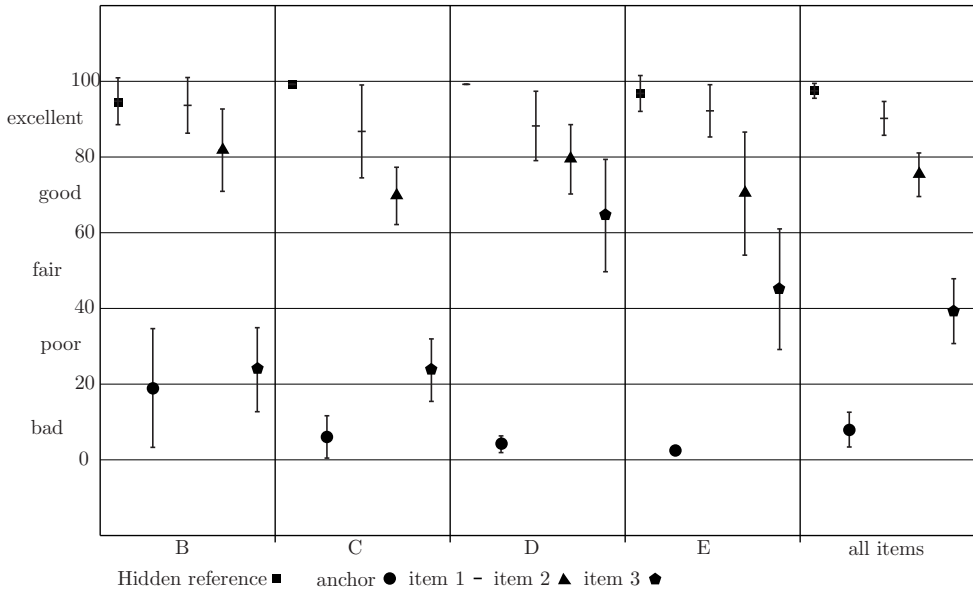
Results and conclusions Figure 8.4(a) and 8.4(b) present the results for the speech and drums example, respectively. In the overall rating the grading depends significantly on the order of reflections which were calculated correctly. While item 1 is in the excellent range and item 2 is still scored as good, item 3 is in the poor range. Considering the different source positions, item 3 is graded poor to fair for all position except D. The reason is the constant direction in this case compared to the reference. Only the distance is changed. The results suggest that the difference in distance of 5 m was not perceived by the subjects. Item 1 and item 2 are always in the excellent to good range. Similar to the first experiment, this experiment supports the proposition that a reproduction of first- and second-order reflections in auralization is sufficient for a high quality auralization. The results of this experiment are reflected in common practice in several auralization and simulation tools (Vorländer, 2008).

8.3 Verification of Single-user WFS Optimization

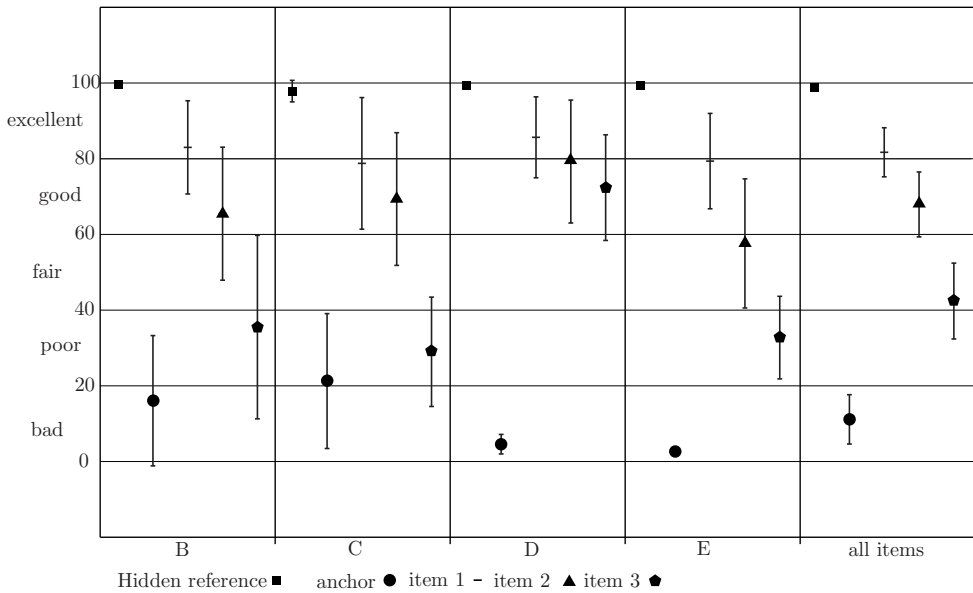
In order to verify the optimization derived in section 7.2.2, two pilot listening experiments were conducted. The experiments took place in an acoustically treated studio at Fraunhofer IDMT. The room is equipped with a 104 channel WFS reproduction system. For the tracking of the user, an infrared tracking system with six cameras was installed, with a measurement accuracy in the order of millimeters.

8.3.1 Optimization for Focused Sources

The first experiment was conducted to verify the optimized loudspeaker selection for focused sources. A focused source is rendered in position $(x, z) = (0, 0.8)$. The source position was indicated with a circle on the floor of the listening area. The circle was marked in ten degree steps. The subjects were positioned at zero degrees at the beginning of the experiment. The task of the subjects was to walk along the circle until they could no longer perceive the source as a focused source. The subjects had to perform the task twice, the optimization was turned on and off in random order. Eleven WFS-experienced subjects participated in this experiment. The sound material was a 90 second piece of pop music. Figure 8.5 shows the result of the experiment. In the case of an untracked user (Figure 8.5(a)), the area in which the source is perceived correctly (gray) is significantly smaller than the theoretically determined area



(a) Speech example.



(b) Drums example.

Figure 8.4: Average results and 95% confidence intervals for the source position experiment.

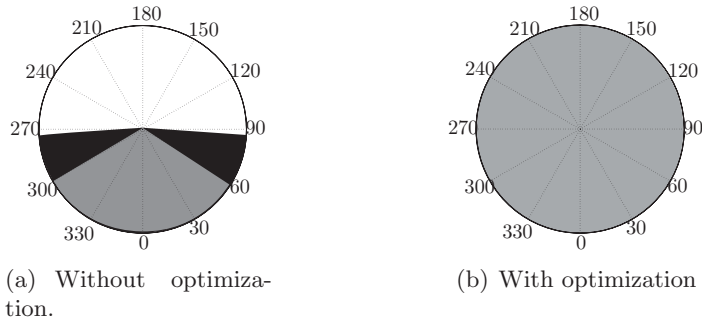


Figure 8.5: Results for the focused source listening experiment. (a) Black: Theoretical reconstruction area of a focused source, Gray: Mean value of the subjective evaluation without optimization. (b) Gray: Mean value of the subjective evaluation with optimization (Sladeczek, 2008).

(black). In case of the optimized sound field (Figure 8.5(b)), the focused source can be perceived from any angle around the source. Due to the fact that the speaker system is visible during the test, some subjects indicated that the perception of the source was easier if they closed their eyes and moved their head closer to the source position.

8.3.2 Diffuse Field Reproduction

In the second listening experiment the enhancement of diffuse field reproduction due to optimization of the level and delay for a limited number of plane waves was verified. The task for the subjects was to indicate on a five point scale whether they were able to perceive a dominant direction in the sound field or not. In the case of a perception of a dominant direction, the direction was indicated by the subjects. The sound field consisted of eight plane waves as indicated in Figure 8.6. The signals used were eight uncorrelated pink noise signals played back simultaneously. First the subjects were trained. Positioned at position *C*, the subjects had to listen to a diffuse reference in comparison to a stimulus with 5 dB attenuation for a plane wave of one direction. The training was performed until the subjects gave reliable results when indicating the direction of the attenuation. During the experiment each subject was at positions *A*, *B*, *C* and *D*. Each position was tested twice, once with the optimization active and once without optimization. The sequence of positions and stimuli were randomized. Eleven WFS-experienced subjects participated

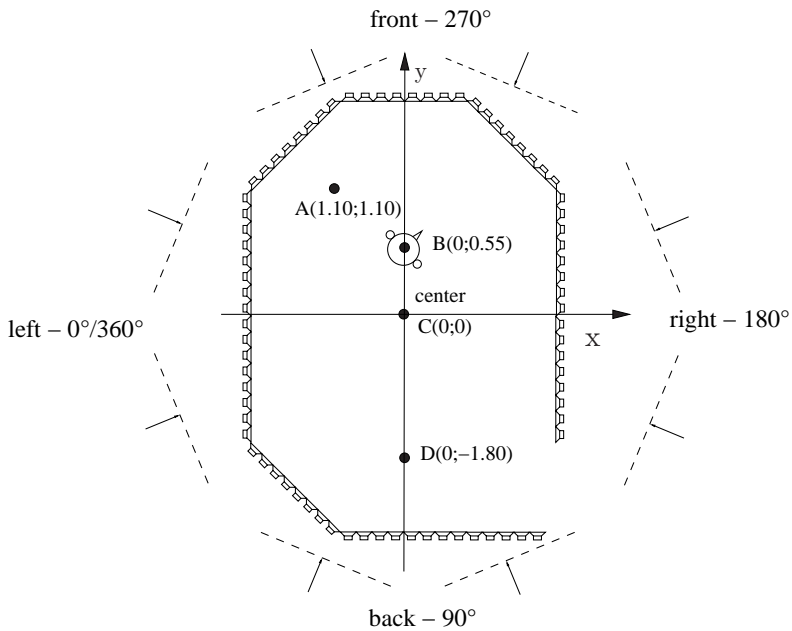


Figure 8.6: Geometry of the diffuse field experiment. The four listener positions are marked A , B , C and D . The directions of incidence for the eight plane waves are indicated as well as the used loudspeaker set-up (Ślądcezek, 2008).

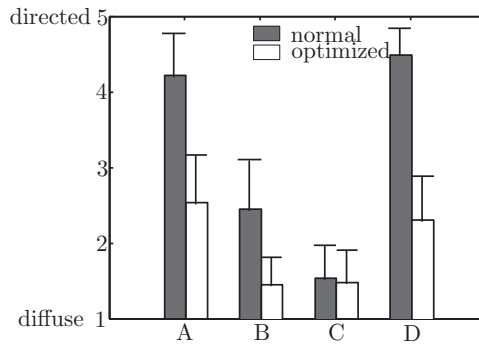


Figure 8.7: Mean values and standard deviation of the diffuse field grading experiment for position *A* to *D* as indicated in Figure 8.6.

in the experiment. Figure 8.7 and 8.8 show the results of the experiment. The optimization leads to an improvement of the perceived diffuseness of the sound field (Figure 8.7). This holds for positions *A*, *B* and *D*. In position *C* no optimization is achieved. *C* is the original reference position of the WFS system, so this result was expected. The amount of directional impression is correlated to the position (Figure 8.8). As expected, a position with a longer distance to the reference point leads to a bias of the sound field in the corresponding direction. The remaining bias may be caused by the fact that a different number of secondary sources is active for different virtual source positions. This can result in a perceived level difference.

8.4 Perception of Virtual Microphones

This section presents the results of perceptual experiments using simulated virtual microphones based on spherical array apertures. The aim of the listening experiments is to compare the results of the WFE based on spherical sensor aperture measurements in different setups with an ideal impulse response, simulated using the desired virtual stereo microphone setup. The investigation is separated into two experiments. While the first is dedicated to the direct sound and early reflections, the second experiment concentrated on the diffuse sound field. To achieve controlled conditions, both experiments used simulated data.

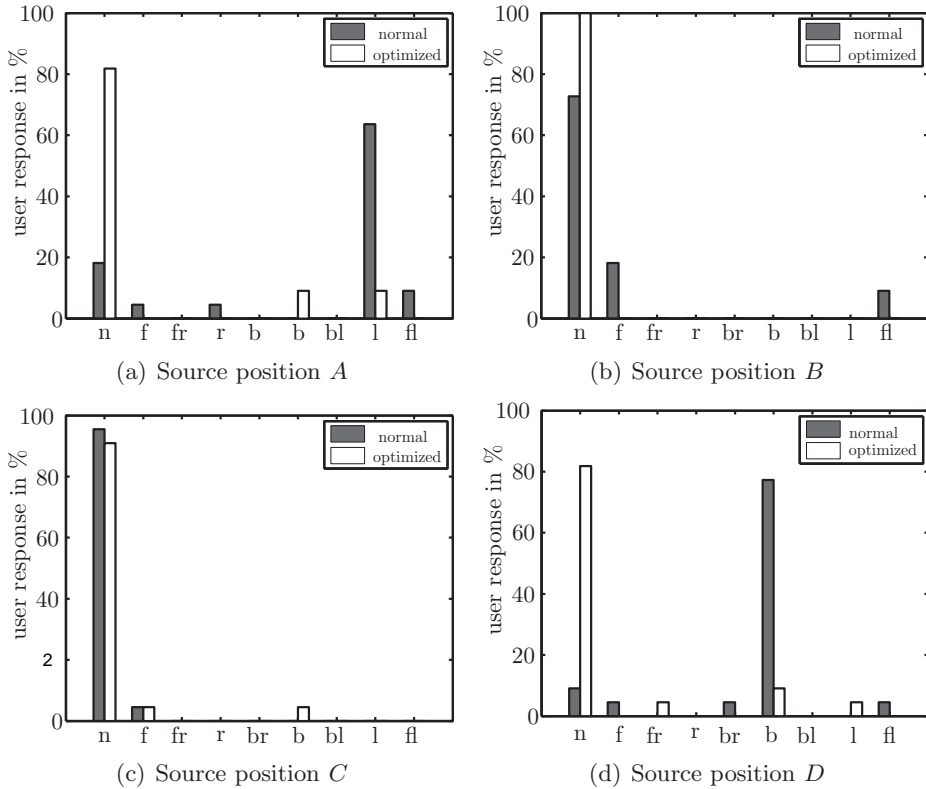


Figure 8.8: Histogram of the indicated directions during the diffuse grading experiment. The positions of the subjects correspond to positions as shown in Figure 8.6 (n - no direction; f - front; r - right; l - left; b - back) (Sladeczek, 2008).

8.4.1 Listening Experiment with Mirror Image Source Model

Experiment Design

For the purpose of this listening experiment, a mirror image source simulation was calculated. The layout of the room is given in Figure 8.9(a). The reflection coefficients of the walls are frequency-independent, resulting in an overall reverberation time of 0.5 s. The model does not include any diffuse energy. Only distance-dependent attenuation and air damping were simulated for each reflection. Based on this simulation, three different virtual microphone setup, as given in Figure 8.9(b), were calculated (coincident figure-of-eight BL, ORTF,

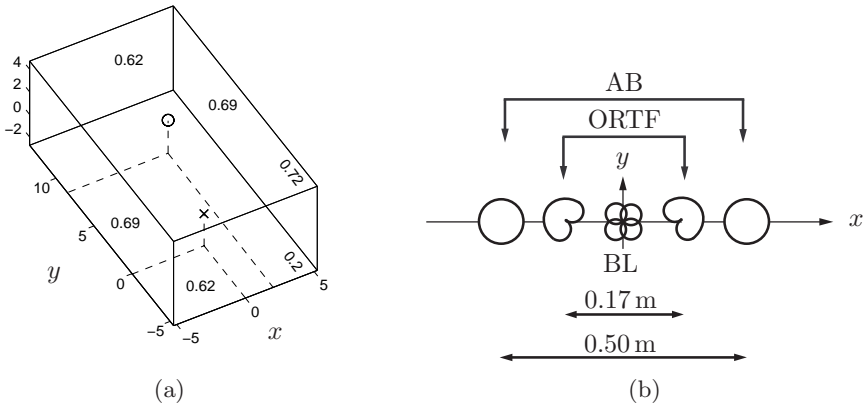


Figure 8.9: Plot (a) shows the geometry of the room used for the MISM simulation with reflection coefficients for each wall, the source position (o), and the receiver position (x). Plot (b) shows the geometrical layout of the simulated virtual microphone setups.

and AB with 50 cm distance). The resulting impulse responses were convolved with dry audio files consisting of pink noise bursts (i.e., 50 ms burst followed by a 2 s pause), pink noise, male speech, and music (drums and bass guitar). In addition to ideal simulation of the three stereophonic setup, six different spherical arrays, as given in Table 8.4, were simulated. From the output of these arrays, the three stereophonic setups are computed using the spherical harmonic extrapolation method described in Section 5.3. For the dual radius arrays 2 and 6, the transition frequency f_c was chosen such that it corresponds with the aliasing frequency of the large radius r_2 . This means that from a physical point of view, both arrays are designed with optimal parameters (i.e., parameters which ensure low aliasing). This is in contrast to the dual radius arrays 4 and 5, for which the transition frequency was set above the aliasing frequency. To avoid artifacts and ensure the highest possible resolution a radius switching was applied utilizing a notch filter with a width of 60 Hz at the transition frequency for arrays 4 and 6. In the listening experiment, a hidden reference and an anchor consisting of the dry audio signal were included. The test was performed using electrostatic headphones and high quality A/D converters. Before the three test sessions, the subjects had the opportunity to listen to the test items in order to familiarize themselves with the audio material and the expected artifacts. 15 subjects participated in the listening experiment (Two female and 13 male with an age between 23 and 35).

Array	r_1	r_2	f_c	Position error	Noise	Filter
1	0.28 m	-	-	none	none	none
2	0.28 m	1 m	2.1 kHz	none	none	none
3	0.28 m	-	-	1°/4°	-80 dB	none
4	0.28 m	1 m	3.15 kHz	1°/4°	-80 dB	3.15 kHz
5	0.28 m	1 m	3.15 kHz	none	none	none
6	0.28 m	1 m	2.1 kHz	1°/4°	-80 dB	2.1 kHz

Table 8.4: Open sphere array configurations used for the listening experiment. Arrays 4 and 6 utilize a 60 Hz wide notch filter to suppress the sinusoid at the transition frequency of the two radii. A realistic directivity was included in configurations 3, 4, and 6. For all configurations Q_{Leb} was set to 2030 and $f_{\max} = 20$ kHz.

Experimental Results

Figure 8.10 presents the results for the spaced omni-directional configuration (AB). In the overall rating (*all items*), the dual radius ideal setups^(2,5) are scored significantly better than the single radius ideal setup⁽¹⁾. This implies that in the latter setup, artifacts due to the small area of accurate extrapolation (AAE) and the large extrapolation range become audible. The single radius real setup⁽³⁾ scored significantly worse than the dual radius real setups^(4,6) and dual radius ideal setup⁽²⁾, for the same reason. The dual radius ideal setup^(2,5) scored significantly better than the dual radius real setups^(4,6), as was expected. In case of the noise bursts, the single radius setups^(1,3) scored significantly worse than the dual radius setups^(2,4,5,6). For the drums and noise examples, the ideal setups^(1,2,5) are graded significantly better than real setups^(3,4,6). For the speech example, the dual radius ideal setups^(2,4) are graded better than the dual radius real setups^(4,6), which are even better than the single radius setups^(1,3). The observed effects can be explained by the large extrapolation range of the virtual stereo microphone setup which is close to the radius of the small single radius setup. Figure 8.11 presents the results for the coincident figure-of-eight configuration. No significant differences can be found between the different setups in the overall rating. The reason is that all VMs are located inside the AAE, even when using the small single radius setup. Nonetheless, for the noise bursts the real setup tend to be score higher than the ideal setup. An explanation is that the spatial aliasing error is highly concentrated in specific areas when simulating ideal sensor arrays. These errors can be perceived very clearly. In the real setup, the aliasing error is spread over

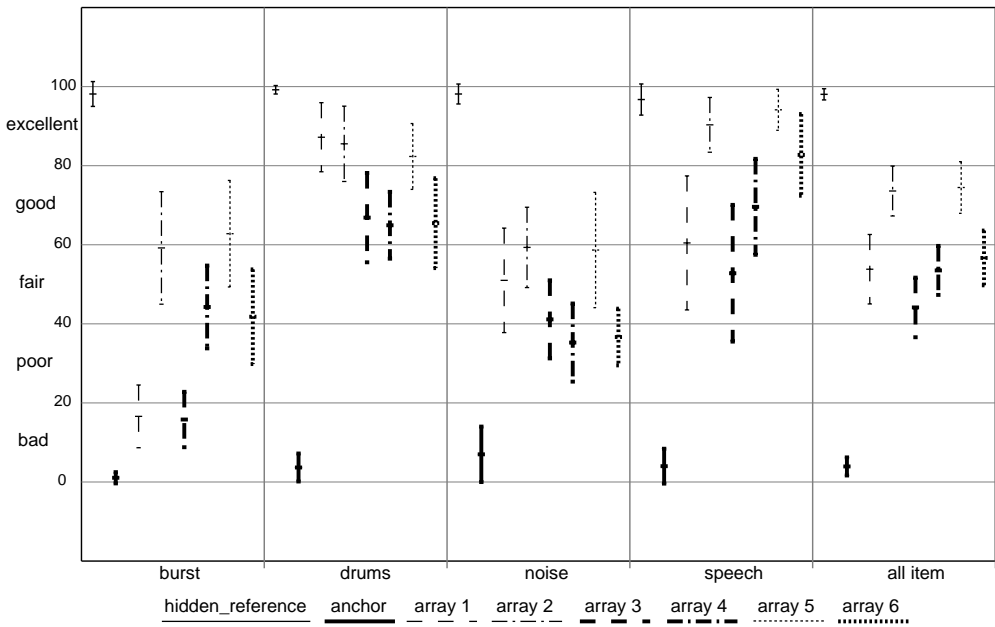


Figure 8.10: Average results and 95% confidence intervals of the mirror image source listening test for the AB set-up.

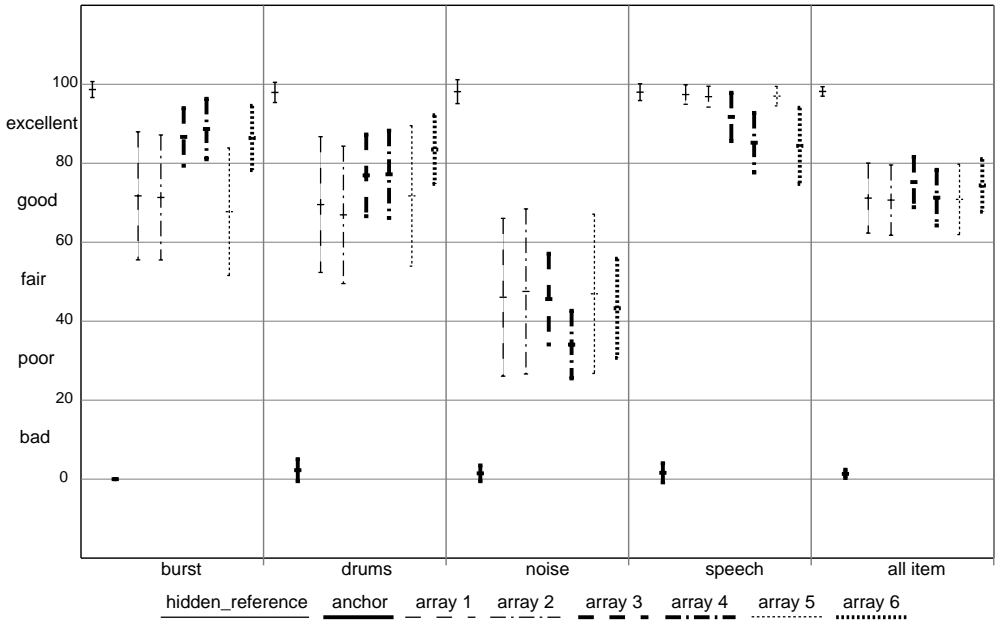


Figure 8.11: Average results and 95% confidence intervals of the mirror image source listening test for the coincident figure-of-eight setup (BL).

a wider spatial area and is thus less audible. Figure 8.12 presents the results for the ORTF configuration. In the overall rating both dual radius real setup tended to be rated better than all single radius setups and the dual radius ideal setup. This is a consequence of the strong artifacts when the extrapolation is close to the edge of the small radius. The tendency to grade the real setups better than the ideal setup for the noise bursts and drums is due to spatial aliasing, as described for the results of the BL setup.

8.4.2 Listening Experiment Diffuse Field

Experiment Design

For the purpose of this listening experiment, a diffuse field simulation was calculated without any reflections and direct sound. The aim of the experiment was to compare the virtual microphones, which were generated based on spherical array measurements, with ideal microphone setups calculated based on the same diffuse field. Furthermore, a different approach to calculate the

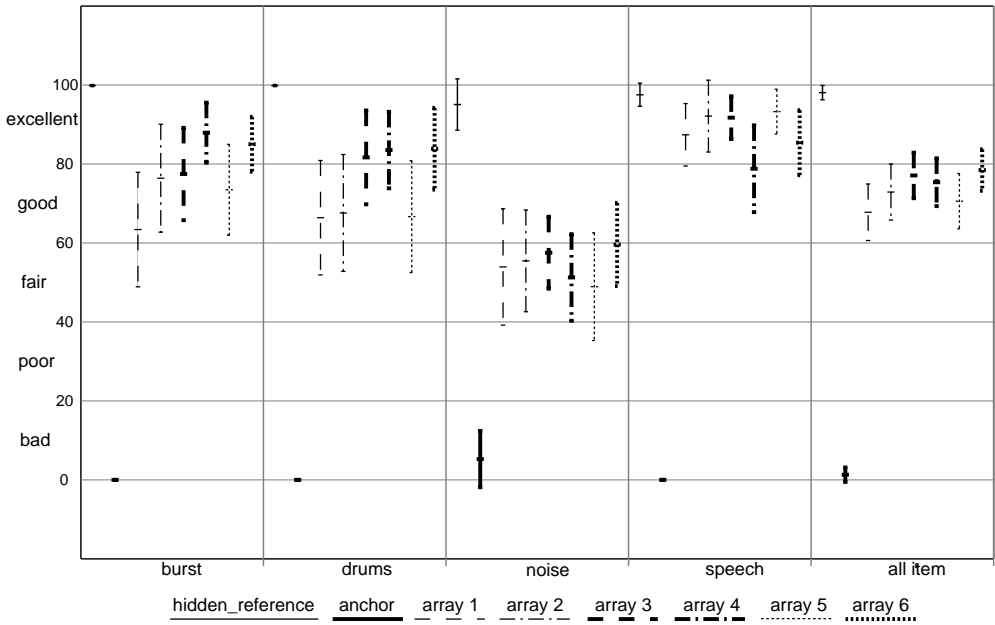


Figure 8.12: Average results and 95% confidence intervals of the mirror image source listening test for ORTF setup.

required impulse responses is included. This approach used the decorrelation techniques described in Section 2.7.2. The stereo configurations used are given in Figure 8.9(b). The diffuse sound field was simulated by calculating a plane wave sound field with 1202 directions using a Lebedev grid. For each plane wave a decorrelated white noise signal was generated. This was temporally shaped using an envelope. The parameters used for the envelope are a pre-delay of 5 ms, an attack time of 30 ms, a sustain time of 15 ms and a release time of 500 ms. An air-damping filter was applied using a STFT time variant filter based on the mean distance. In addition, an ideal simulation of the three stereophonic setups, six different spherical arrays, as given in Table 8.4, were simulated. From the output of these arrays, the three stereophonic setup are computed using the spherical harmonic extrapolation method described in Section 5.3. Furthermore, a simulation of the desired coherence function was included in the experiment. For this purpose the impulse response of the left reference was used. Based on this, a second impulse response was derived using the method described in Section 6.3.2. The coherence function required was calculated using Eq. (A.43). In the listening experiment, a hidden reference and an anchor consisting of a mono version of the reference audio signal were included. The test was performed using electrostatic headphones and high-quality A/D converters. Before the three test sessions, the subjects had the opportunity to listen to the test items in order to familiarize themselves with the audio material and the expected artifacts. 15 subjects in the age range of 22 to 43 years participated in this listening experiment. No hearing problems were reported by the three female and 12 male subjects.

Experimental Results

For the AB set-up the results are shown in Figure 8.13. It can be seen that, in the overall rating, the dual radius setups without error ^(2,5) were graded significantly better than the single radius setup without error ⁽¹⁾. The simulation scored in the upper range between the dual radius and the single radius. The dual radius setup including errors ⁽⁶⁾ was graded slightly better than the single radius including errors ⁽³⁾. In view of the results for the different audio signals, it was found that when using the speech signal, nearly all arrays yield excellent results, only the simulation was slightly worse. The reason is that the spectral content of the speech signal excludes the critical frequency range of the arrays. In case of the noise stimulus the ratings tend to be worse because spectral differences were perceived more easily by all of the subjects. The sin-

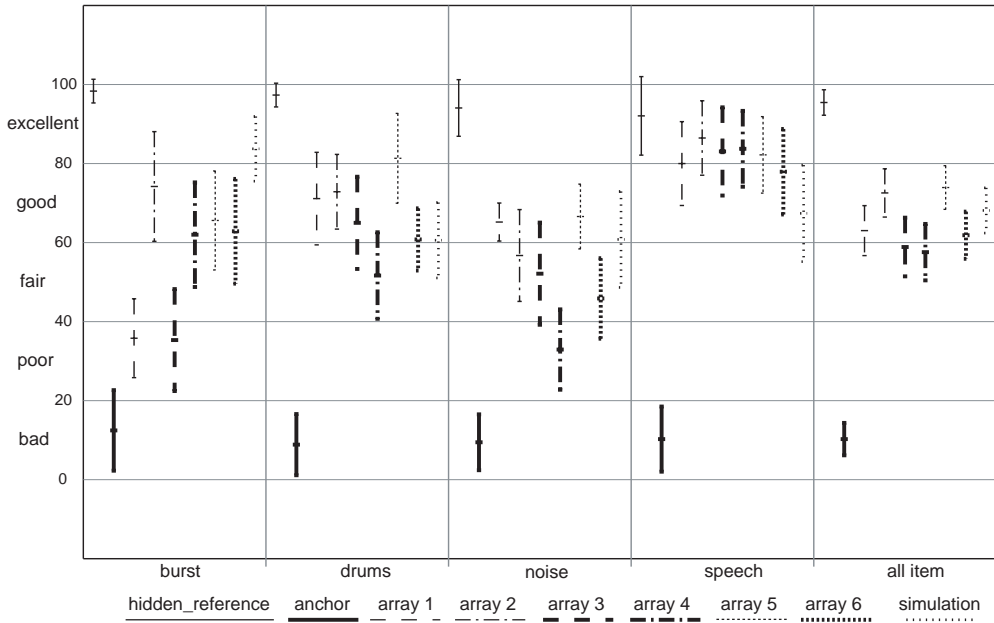


Figure 8.13: Average results and 95% confidence intervals of the diffuse field listening experiment for AB set-up.

gle radius error free array ⁽¹⁾ was graded much better for this signal than in the overall results. For the noise bursts the single radius setups ^(1,3) were graded significantly worse than all other setups ^(2,4,5,6). In this setup the extrapolation error of the single array radius design is much stronger than in case the of the dual radius and the simulation configuration. Furthermore, the extrapolation error can be perceived much better in transient signals than in static noise. In the case of the noise burst, the dual radius setup as well as the simulation performed much better than the single radius setup. The results for the ORTF setup are presented in Figure 8.14. In the overall results the different setups are very comparable. The simulation is in the same range as the optimal dual radius setup ⁽⁶⁾. In comparison to the AB setup, the single radius setups ^(1,3) performed much better. The reason is the smaller extrapolation range in this configuration. In the BL setup the difference between arrays including errors ^(3,4,6) and setups without ^(1,2,5) can be distinguished. The simulation is in the same range as the error-free setups. In comparison to the mirror image source test, the results are not that clear because the smoothing of the aliasing error due to the measurement errors cannot be perceived clearly in a sound field that is only diffuse. One can conclude again that the audio signal has very strong

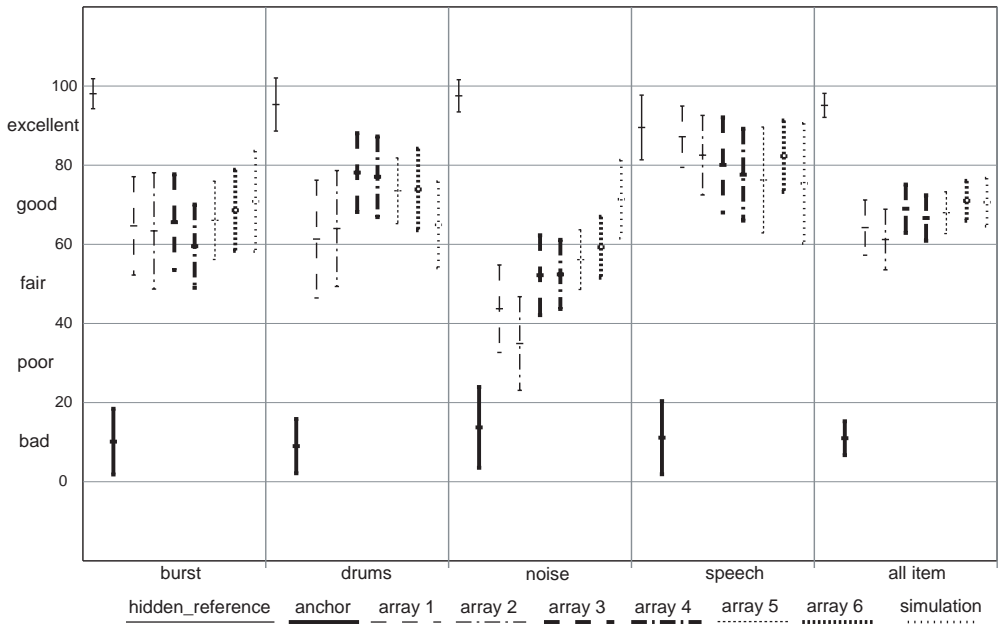


Figure 8.14: Average results and 95% confidence intervals of the diffuse field listening experiment for the ORTF setup.

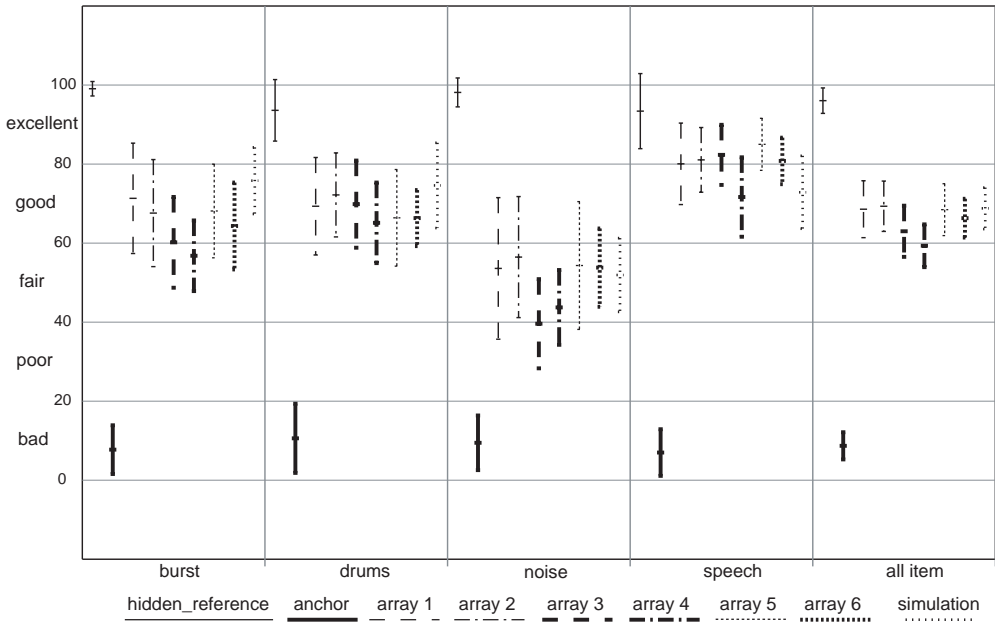


Figure 8.15: Average results and 95% confidence intervals of the diffuse field listening experiment for the coincident figure-of-eight setup (BL).

influence on the perception of different artifacts of the array processing. In the case of diffuse sound fields for virtual microphone setups, the simulation of the coherence function yields results comparable to those of the array processing. This is important because the computational effort is very high for complete array processing with impulse responses. A possible solution is to use the complete array processing only for the early part of the impulse response and calculate the late part using the simulation method. In summary it was shown that the dual radius spherical array including realistic measurement errors can generate an auralization quality close to the ideal simulation and the error-free case.

8.5 Perception of Binaural Reproduction

To study the adaptation of spherical array room impulse response measurements to binaural reproduction, a listening test was performed. The influence of different array designs and measurement errors was investigated. Seven array setups were studied as defined in Table 8.5.

Array	r_1	r_2	f_c	Position error	Noise	Processing
1	0.28 m	-	-	none	none	none
2	0.28 m	-	-	$1^\circ/4^\circ$	-80 dB	none
3	0.28 m	1 m	2.1 kHz	none	none	Switch
4	0.28 m	1 m	2.1 kHz	$1^\circ/4^\circ$	-80 dB	Switch/Notch
5	0.28 m	1 m	2.1 kHz	none	none	Fade/Adapt
6	0.28 m	1 m	2.1 kHz	$1^\circ/4^\circ$	-80 dB	Fade/Adapt
7	0.28 m	1 m	2.1 kHz	$1^\circ/4^\circ$	-80 dB	Switch

Table 8.5: Open sphere array configurations used for the binaural listening experiment. Array 4 utilizes a 60 Hz wide notch filter to suppress the sinusoid at the transition frequency of the two radii. Array 5 and 6 uses a fading of the coefficients and an adaptation of the resolution at the transition frequency. A realistic directivity was included in configurations 2, 4, 6, and 7. The position error was normally distributed and is given in azimuth and elevation. Furthermore, a 3° offset error was included in the co-elevation. For all configurations Q_{Leb} was set to 2030 and $f_{\max} = 20$ kHz.

8.5.1 General Design of the Listening Experiments

In all parts of the listening experiments, the array configurations given in Table 8.5 were used. In each case the anchor was a single channel signal generated as a sum of the two channels of the reference signal, for which a diotic¹ reproduction was used. The experiment was a grading experiment according to (ITU, 2003). Four audio signals were used: speech, drums and bass, pink noise, and 50 ms pink noise bursts. They were chosen because the different effects or artifacts of the array processing needed to be analyzed. The pink noise signal is optimal to grade the difference in coloration. The bursts are optimal to identify short spectral effects such as ringing at specific frequencies. The speech and the drums and bass examples were chosen as real-life signals, they have strong transients and all listeners are very familiar with speech.

8.5.2 Listening Experiment Mirror Image Source Model

Experiment Design

In the first experiment the array configuration was used in a simulation based on the same MISM as was used and illustrated in Section 8.4.1 (see Figure 8.9(a)). The reference was generated by convolving all discrete reflections with the HRIRs from the high resolution measurement corresponding to their angle of incidence. The reflection positions were rounded to the nearest position in a Chebyshev grid with a two degree resolution in azimuth and elevation. 24 experienced subjects participated in the listening experiment. The results of three subjects were omitted because their grading of the anchor was better than poor in all examples.

Experimental Results

The results of the experiment are depicted in Figure 8.16. In the overall results it can be noticed that the single radius setups^(1,2) were graded significantly worse than the dual radius setups^(3,4,5,6). The only exception is the array ⁽⁷⁾

¹The following three categories of signal presentation using headphones are often distinguished: Monotic presentation, in which a signal is present in only one headphone, diotic presentation, in which the same signal is present in both headphones, and dichotic presentation, in which different signals are present in each headphones (Blauert, 1997).

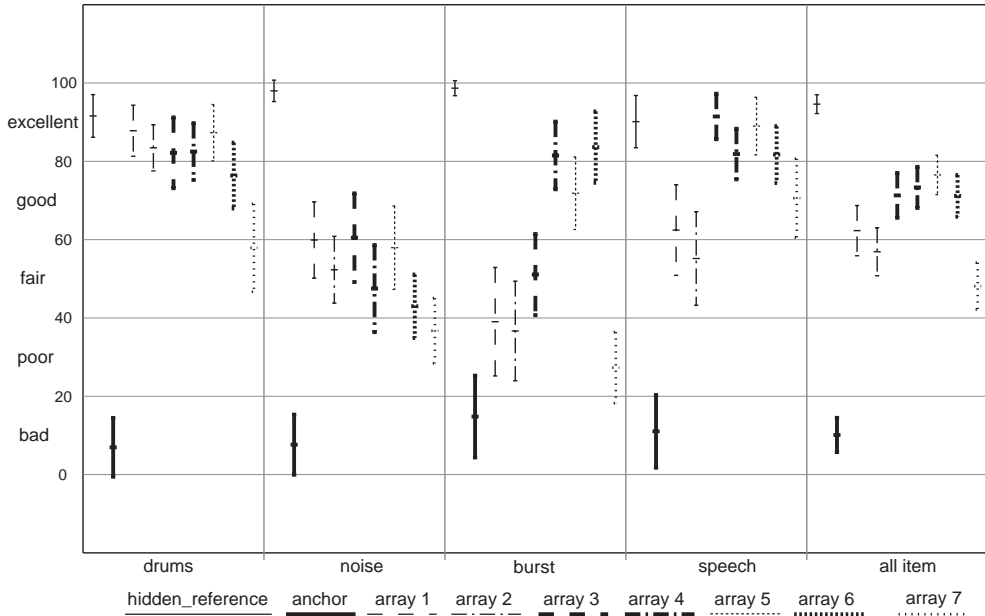


Figure 8.16: Average results and 95% confidence intervals for the binaural mirror image source model simulation experiment.

because of its ringing at the transition frequency, which was clearly audible in the case of drums, pink noise, and especially for the pink noise bursts. For the drums example nearly no difference between the different array designs is noticeable. For the pink noise signal the coloration due to aliasing becomes clearly audible and leads to a significant difference to the reference grading for all arrays. In case of the pink noise burst, only the arrays without ringing artifacts^(4, 5, 6) were graded from excellent to good. In summary, it can be stated that the dual radius configuration clearly outperforms the single radius design used, but care must be taken for the adaptation at the transition frequency to avoid artifacts. It is interesting that the adaptation and fading used in arrays^(5, 6) is not significantly better than the notch filter used in array⁽⁴⁾.

8.5.3 Listening Experiment Diffuse-field Simulation

Experiment Design

In the second part of the listening experiment a diffuse-field simulation was used. The simulation was the same as was used for the experiment in Section 8.4.1. The reference was generated by convolving all of the plane waves of the diffuse field simulation with an appropriate HRIR. The anchor again was a diotic presentation of a sum of the two reference channels. 24 subjects participated in the experiment, and again three subjects were omitted because they did not identify the anchor.

Experimental Results

The results of the diffuse field experiment can be found in Figure 8.17. The results are nearly identical to the results of the mirror image source model experiment. Again the dual-radius designs without ringing artifact^(3, 4, 5, 6) lead to significantly better results than the single radius setups^(1, 2). The gradings for the different audio examples are similar. It is interesting to notice that for the noise example, there is a tendency to grade the ideal setups^(1, 3, 5) better than the non-ideal setups including measurement errors^(2, 4, 6, 7).

8.5.4 Listening Experiment Using a Plane Wave Simulation

The fact that the notch filter approach used in array design⁽⁴⁾ leads to nearly identical results as the adaptation-and-fade approach used in array designs^(5, 6) was surprising. In an informal pre-test the adaptation approach led to subjectively better results. One assumption was that the complex diffuse field, as well as the room impulse response, masked some artifacts, which still exist when using the notch filter in design⁽⁴⁾. For this reason a third experiment was performed. Here, one single plane wave was simulated. As a reference, the audio signal was only convolved with the HRIR of the DOA of the plane wave. The anchor was again a diotic presentation of the sum of the two channels of the reference signal. The same subjects as in the previous experiment participated. Four subjects were omitted in the results because they did not identify the anchor.

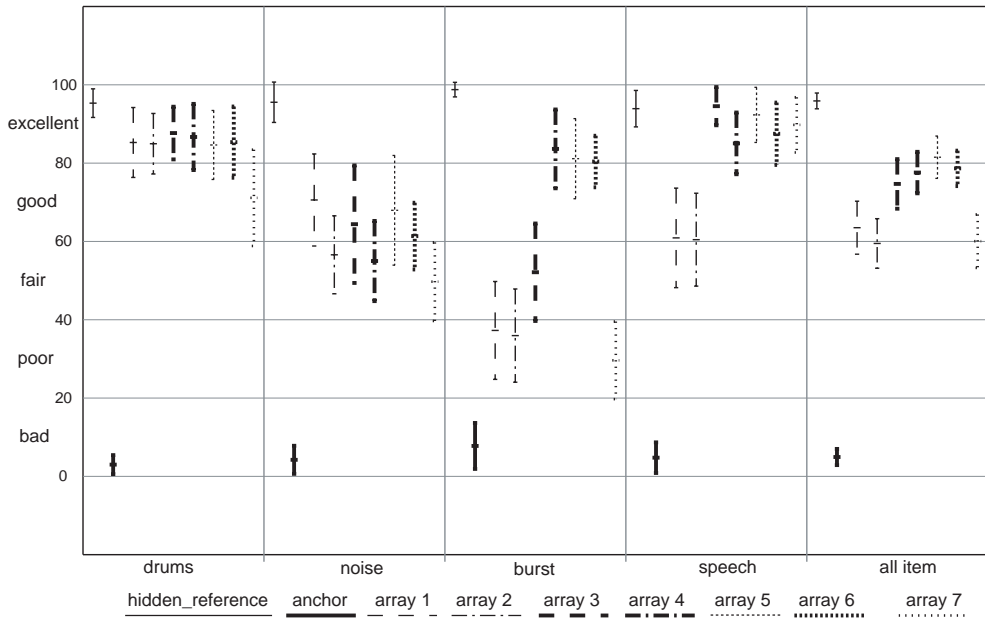


Figure 8.17: Average results and 95% confidence intervals for the binaural diffuse field simulation experiment.

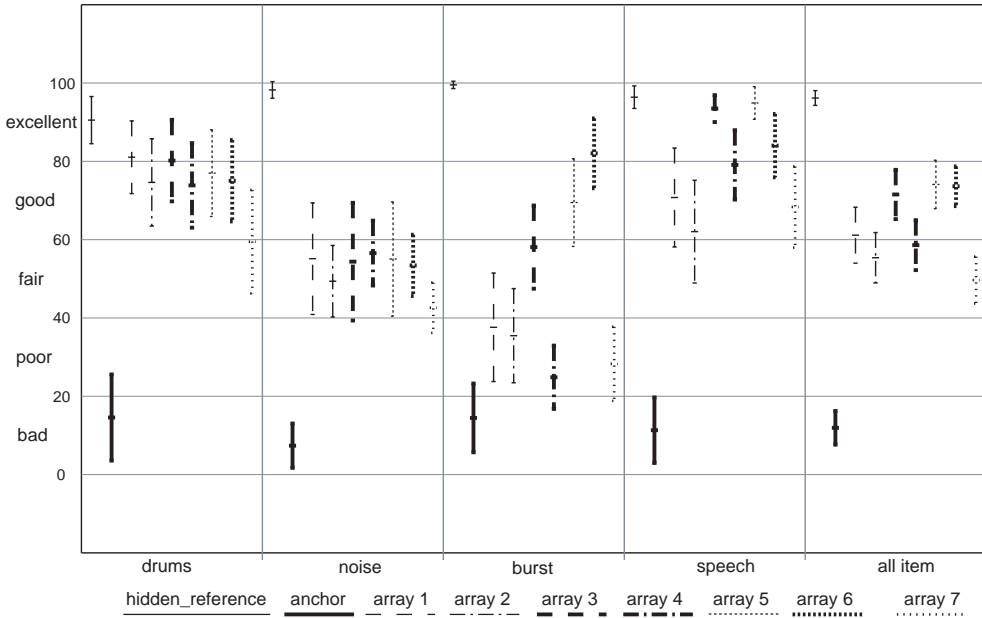


Figure 8.18: Average results and 95% confidence intervals for the binaural plane wave simulation experiment.

Experimental Results

The results (see Figure 8.18) now show a difference in the grading for array ⁽⁴⁾. While the results for the different arrays and audio examples are comparable to the previous experiments, the notch filter design was graded significantly worse. The reason is that although the filter suppresses the sinusoid component, in transient signals a short ringing can be noticed. This also holds for the speech and for the pink noise burst signals. The drums example is also very transient, but the effect is masked due to the continuous signal. The same was true for the noise signal.

8.6 Summary

The results of several perceptual evaluations were presented in this chapter. It was shown that for the binaural simulation of a mirror image source model, the modeling of first- and second-order reflection is sufficient to simulate the source position and direction in case of a directive source with high quality grading compared to an open reference. To generalize the result, a comprehensive study is required, which is beyond the scope of this thesis. Nevertheless, the results conform the literature, stating that a low number of first- or second-order reflections are sufficient to simulate the key parameter (position and direction) of source inside a room. For WFS adaptation and reproduction, a single-user optimized WFS system was perceptually evaluated. The system optimized the reproduction of focused sources and diffuse fields for a single listener. It was shown that it is possible to extend the area of correct reproduction for a focused source to 360 degrees for a single tracked listener. Furthermore, it was shown that a single-user optimized WFS system can enhance the quality of a diffuse field reproduction significantly using a level and delay optimization for a limited number of plane waves. For stereophonic reproduction, the reproduction quality of diffuse field and mirror image source simulations was studied using a quality grading experiment. In this experiment an ideal stereophonic simulation using different microphone setups was compared to virtual microphone signals generated out of a simulated spherical array measurement. It was shown that the dual radius spherical array including realistic measurement errors can generate a high-quality auralization close to the ideal simulation and the error-free case. In case of binaural reproduction, different spherical array designs were compared in terms of auralization quality using binaural reproduction in a listening experiment. It was shown that the dual radius design has advantages in terms of quality in comparison to the single radius design. In the design process, care has to be taken in the correct adaptation of the different frequency ranges for the different subarrays. It was shown that in case of binaural reproduction the spatial resolution adaptation leads to the best results. All listening experiments in this section were performed with simulated data. This was to make sure that all the setups vary only in the parameter under investigation. To generalize the results, these experiments should be performed with different simulations while altering the acoustic properties of the simulated environments.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

A novel processing chain for spatial sound design based on measured room impulse responses has been developed. The system is based on spherical array measurements of room impulse responses. New interaction methods for sound designers have been developed based on such measurements. The interaction principles developed are independent of specific reproduction systems and based on direct interaction with visualizations of spatial impulse responses. Following this processing chain, each building block has been analyzed in detail.

The properties of cardioid open sphere virtual microphone arrays were investigated in terms of error robustness and spatial sampling. The capabilities of extrapolation and plane wave decomposition of such systems were analyzed. The combination of simultaneous measurements on different radii to extend the usable frequency range was investigated and implemented. The extraction of single events (e.g., reflections) from spherical array measurements was studied by measurements in an anechoic room equipped with a single reflecting surface. It was shown that with adequate spatio-temporal filtering, the frequency response can be extracted. The required analysis of measured room impulse responses has been described, and a storage format for the analyzed spatial impulse response has been developed and used in the realized processing framework.

Based on the analyzed impulse responses, a taxonomy of impulse response visualization was given. Suitable interaction techniques for the direct interaction with visualized direction-dependent impulse responses were developed. The interaction process was classified in static and dynamic interactions. For the static interaction, which is based on modifications independent of source

parameter, new methods like inverse energy decay curve editing and the concept of shaping surfaces were developed and applied. Time-variant filtering based on the short-time Fourier transform and the spatial envelope are new principles studied in this context. The dynamic interaction was analyzed, and methods for an efficient design of spatial acoustic scenes were developed. The interaction techniques that are proposed have been realized within a graphical user interface for desktop use. An extension of a new user interface has been described and prototypes have been realized using an augmented reality framework and state-of-the-art user interface hardware.

The dynamic interaction was studied for wave field synthesis, stereophonic reproduction, and binaural reproduction. Methods for the auralization and adaptation of measured and modified high resolution data to these three reproduction system were developed. The effects of measurement errors on the reproduction quality was investigated using listening experiments employing simulated measurements based on mirror image source models and diffuse-field simulations. The advantages of dual radius cardioid spherical microphone arrays and the developed adaptation methods for stereo, binaural, and WFS reproduction have been demonstrated.

9.2 Future Work

This research provides a starting point for a large variety of future research. The measurement time of the system is very high for the required broadband auralization application. For this reason an investigation of non-uniform resolution could be a possible way to optimize the measurement time and effort. A general issue arose during the work: if a design criterion based on psychoacoustic principles can be found, a psycho-acoustic model can be developed to predict the expected deterioration of the array performance due to measurement errors. This would be an important step for the future development of array processing and spatial audio reproduction systems.

In the field of interaction, the complete implementation of the proposed new user interface techniques in standard production environments is desirable. User studies have to be performed to develop the proposals in detail. To realize fully the proposed concept, a tracking system is required, which has been up to now very expensive. Future developments in this field can lead to an efficient realization of the proposed system.

The methods proposed and developed in this thesis are based on the measurement of impulse responses. Due to a priori knowledge it is possible to analyze and extract several properties of such measurements and allow the interaction modification by a sound designer. The next step is the application of the proposed method to array recordings of complete performances. To make this possible, a multi-channel array recording and analysis technique has to be developed. The author believes that a tempo-spatial editing of recorded events, comparable to the possibilities of visual post-production, will be possible in the future.

9.3 Applications of Results

Major parts of this work, the Spatial Audio Workstation and its room simulation module have been integrated into products of the Fraunhofer Institute Digital Media Technology (IDMT) and the IOSONO GmbH. Furthermore, the dynamic interaction processing has been implemented in the spatial rendering framework of Fraunhofer IDMT used in industrial projects. For the processing of spherical microphone data, a MATLAB toolbox was developed.

Appendix A

Mathematic Formulations

A.1 Coordinate Systems

Depending on the specific problem, different coordinate systems are used. This section presents an overview of the different coordinate systems.

Cartesian coordinate system The position vector \mathbf{x} in Cartesian coordinates is defined by

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (\text{A.1})$$

The wave vector is defined by

$$\mathbf{k} = \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix}. \quad (\text{A.2})$$

A volume element used for integration is given by

$$dV = dx dy dz \quad (\text{A.3})$$

Spherical coordinate system The position vector in spherical coordinates \mathbf{x} is defined by

$$\mathbf{x} = \begin{pmatrix} \phi \\ \theta \\ r \end{pmatrix}. \quad (\text{A.4})$$

The wave vector is defined by

$$\mathbf{k} = \begin{pmatrix} \phi \\ \theta \\ k \end{pmatrix}. \quad (\text{A.5})$$

The relation between spherical coordinates and Cartesian coordinates is given by

$$x = r \sin \theta \cos \phi \quad (\text{A.6a})$$

$$y = r \cos \theta \sin \phi \quad (\text{A.6b})$$

$$z = r \cos \theta \quad (\text{A.6c})$$

$$\phi = \arctan \frac{y}{x} \quad (\text{A.7a})$$

$$\theta = \arccos \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right) \quad (\text{A.7b})$$

$$r = \sqrt{x^2 + y^2 + z^2}, \quad (\text{A.7c})$$

where ϕ, θ, r denote the azimuth, elevation and radius. The parameter range of these variables is $0 \leq \phi \leq 2\pi$, $0 \leq \theta \leq \pi$ and $0 \leq r \leq \infty$. The relation between the wave vector in spherical coordinates \mathbf{k}_H and in Cartesian coordinates \mathbf{k}_C is analog to the position vector.

A.2 Fourier transformations

A.2.1 Fourier Transform

The Fourier transform (FT) $P(\omega) = \mathcal{F}\{p(t)\}$ of a time-dependent signal $p(t)$ is given by (Kress and Irmer, 1989):

$$P(\omega) = \int_{-\infty}^{\infty} p(t) e^{-i\omega t} dt. \quad (\text{A.8a})$$

$P(\omega)$ denotes the complex spectrum, $\omega = 2\pi f$ denotes the temporal frequency and t the time. The inverse Fourier transform (IFT) $p(t) = \mathcal{F}^{-1}\{P(\omega)\}$ is

given by:

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega t} d\omega. \quad (\text{A.8b})$$

A.2.2 Multidimensional Fourier Transformation

The decomposition of an acoustic field in its plane wave components can be interpreted as a specialization of multidimensional Fourier analysis of spatio-temporal signals. This section follows the derivation by Spors (2005). In case of a spatio-temporal signal $p(\mathbf{x}, t)$ the Fourier transform is given by

$$\tilde{P}(\mathbf{k}, \omega) = \int_{\Omega \in \mathbb{R}^D} \int_{-\infty}^{\infty} p(\mathbf{x}, t) e^{i\langle \mathbf{k}, \mathbf{x} \rangle - i\omega t} dt d\Omega. \quad (\text{A.9a})$$

The inverse Fourier transform is given by

$$p(\mathbf{x}, t) = \frac{1}{(2\pi)^{D+1}} \int_{K \in \mathbb{R}^D} \int_{-\infty}^{\infty} \tilde{P}(\mathbf{k}, \omega) e^{-i\langle \mathbf{k}, \mathbf{x} \rangle + i\omega t} d\omega dK, \quad (\text{A.9b})$$

where \mathbf{k} denotes the wave vector, D the dimension and \mathbf{x} the position vector. The expression $\langle \mathbf{k}, \mathbf{x} \rangle$ represents the inner product of \mathbf{k} and \mathbf{x} . $d\Omega$ and dK are volume elements of the position and wave vector space, which have to be defined according to the coordinate system used. The temporal part of the transformation is independent of the coordinate system used. As the Fourier transform in the temporal domain, the spatial Fourier transform and its inverse form a complete set of transformations. A signal can be transformed into the wavenumber-frequency domain and back into the space-time domain using the inverse Fourier transformation without loss of any information. The spatio-temporal Fourier integral can be split into a spatial and temporal part. In other words, the spatio-temporal Fourier transformation can be expressed as a temporal Fourier transformation followed by a spatial Fourier transformation. This property of separability as mentioned above leads to the spatial Fourier transformation given by:

$$\tilde{P}(\mathbf{k}, \omega) = \int_{\Omega \in \mathbb{R}^D} P(\mathbf{x}, \omega) e^{i\langle \mathbf{k}, \mathbf{x} \rangle} d\Omega \quad (\text{A.10a})$$

$$P(\mathbf{x}, \omega) = \frac{1}{(2\pi)^D} \int_{K \in \mathbb{R}^D} \tilde{P}(\mathbf{k}, \omega) e^{-i\langle \mathbf{k}, \mathbf{x} \rangle} dK. \quad (\text{A.10b})$$

A.2.3 Fourier Series

A given function $p(\phi)$ on the unit circle in polar coordinates can be represented by a Fourier series (FS) (Teutsch, 2007), (Williams, 1999):

$$p(\phi) = \sum_{\nu=-\infty}^{\infty} p_{\nu} e^{i\nu\phi}. \quad (\text{A.11a})$$

The Fourier coefficients, p_{ν} are given by

$$p_{\nu} = \int_0^{2\pi} p(\phi) e^{-i\nu\phi} d\phi. \quad (\text{A.11b})$$

A.2.4 Discrete Fourier Transform

In case of a sampled signal $p(n)$, the continuous Fourier transform becomes a discrete Fourier transform (DFT) denoted as $\mathcal{DFT}\{p(n)\}$ given by:

$$P(k) = \sum_{n=0}^{N-1} p(n) e^{-i\frac{2\pi k}{N}n}, \quad (\text{A.12a})$$

(Kress and Irmer, 1989) where n is the discrete sample index and k the discrete frequency index. $P(k)$ denotes the discrete spectrum. The inverse discrete Fourier transform IDFT $\mathcal{DFT}^{-1}\{P(k)\}$ is given with

$$p(n) = \frac{1}{N} \sum_{k=0}^{N-1} P(k) e^{i\frac{2\pi k}{N}n} \quad (\text{A.12b})$$

A.3 Special Functions

A.3.1 Legendre Functions

The Legendre functions $P_n^m(x)$ of an arbitrary argument x are briefly discussed in this section. A detailed discussion can be found in (Williams, 1999, p. 186–190). Legendre polynomials $P_n(x)$ are Legendre functions for $m = 0$ specified by Rodrigues' Formula:

$$P_n(x) = \frac{1}{2^n n!} \cdot \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (\text{A.13})$$

Beside the Legendre polynomials, *associated Legendre functions* exist. These functions are characterized by two integer indices $P_n^m(x)$, with $m \neq 0$. In case of positive m , these functions are related to the Legendre polynomials by

$$P_n^m(x) = (-1)^m (1 - x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x). \quad (\text{A.14})$$

To retrieve the associated Legendre functions for negative m , the following expression is used

$$P_n^{-m}(x) = (-1)^m \frac{(n - m)!}{(n + m)!} P_n^m(x), \quad m > 0. \quad (\text{A.15})$$

The main characteristic of the Legendre functions is that they form a complete set of orthogonal functions for each mode m . This is an important property for the spherical harmonics, which are investigated in the subsequent sections.

A.3.2 Spherical Bessel and Hankel Functions

The spherical Bessel functions of the first and second kind, $j_n(x)$ and $y_n(x)$, are related to the corresponding Bessel functions (Collins, 1998, pp. 73–80), denoted by $J_n(x)$ and $Y_n(x)$, as follows (Williams, 1999, p. 194):

$$\begin{aligned} j_n(x) &\equiv \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x), \\ y_n(x) &\equiv \sqrt{\frac{\pi}{2x}} Y_{n+\frac{1}{2}}(x). \end{aligned} \quad (\text{A.16})$$

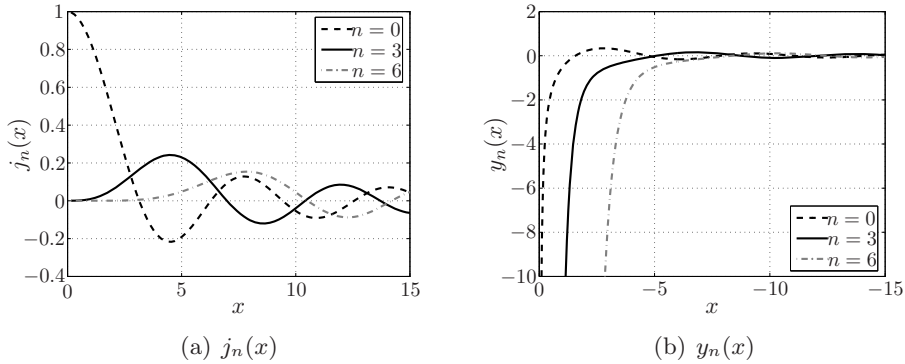


Figure A.1: Spherical Bessel functions of the first kind $j_n(x)$ and the second kind $y_n(x)$ for $n \in \{0, 3, 6\}$ (Thiergart, 2007).

The expressions in (A.16) are valid for $n \in \mathbb{R}$. According to (Williams, 1999, p.194), the spherical Hankel functions of the first and second kind, $h_n^{(1)}(x)$ and $h_n^{(2)}(x)$, are defined as

$$\begin{aligned} h_n^{(1)}(x) &\equiv j_n(x) + i \cdot y_n(x), \\ h_n^{(2)}(x) &\equiv j_n(x) - i \cdot y_n(x). \end{aligned} \tag{A.17}$$

Please note that if x is real, $h_n^{(2)}(x)$ is the complex conjugate of $h_n^{(1)}(x)$. Figure A.1(a) illustrates the spherical Bessel functions of the first kind $j_n(x)$ for different levels n . The functions are finite at the origin, and only $j_0(x)$ is non-zero for that case. The spherical Bessel functions for $n > 0$ show an initial zero value region around the origin, which is larger for higher n . The functions decay rapidly for approximately $n > x$. The spherical Bessel functions of the second kind $y_n(x)$ are depicted in Figure A.1(b), showing a non-finite behavior at the origin. It is clear from Eq. (A.17) that the spherical Hankel functions of the first and second kind exhibit a singularity at $x = 0$ as well.

A.3.3 Spherical Harmonics

The definition of the spherical harmonic functions $Y_n^m(\varphi, \vartheta)$ is given in Eq. (2.34) in a spherical coordinate system. (φ, ϑ) and (φ_0, ϑ_0) denote two different directions separated by the angle θ . In view of Eq. (A.15), it is obvious that a spherical harmonic for negative m can be retrieved from the solution for

positive m by

$$Y_n^{-m}(\varphi, \vartheta) = (-1)^m \overline{Y_n^m(\varphi, \vartheta)}, \quad m > 0, \quad (\text{A.18})$$

where $\overline{Y_n^m(\varphi, \vartheta)}$ is the complex conjugate of $Y_n^m(\varphi, \vartheta)$. Since $-n \leq m \leq n$, there exist $2n + 1$ different spherical harmonics for each level n . Compared to the Legendre functions, the spherical harmonics are not only orthogonal but also orthonormal (Williams, 1999, p.191). Hence,

$$\int_{\mathcal{S}^2} Y_n^m(\varphi, \vartheta) \overline{Y_{n'}^{m'}(\varphi, \vartheta)} \, d\Omega = \delta_{n'n} \delta_{m'm}, \quad (\text{A.19})$$

where $\delta_{n'n}$ is the Kronecker delta which is 1 if n' and n are equal, and 0 otherwise. The integral over the unit sphere \mathcal{S}^2 in Eq. (A.19) is given by definition

$$\int_{\mathcal{S}^2} d\Omega \equiv \int_0^{2\pi} d\varphi \int_0^\pi \sin \vartheta \, d\vartheta. \quad (\text{A.20})$$

The spherical harmonics play a crucial role throughout the 3D analysis of acoustic wave fields, as any arbitrary square integrable function $f(\theta, \phi)$ on a sphere can be decomposed into a sum of spherical harmonics. This is also known as inverse spherical Fourier transform ISFT (Driscoll and Healy Jr., 1994, p.202) and can be written as

$$f(\varphi, \vartheta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n F_{nm} Y_n^m(\varphi, \vartheta), \quad (\text{A.21})$$

where the complex constants $F_{nm}(k)$ are the Fourier coefficients. Since the spherical harmonic functions are orthonormal, the (forward) spherical Fourier transform SFT to determine the coefficients F_{nm} can be computed by

$$F_{nm} = \int_{\mathcal{S}^2} \overline{Y_n^m(\varphi, \vartheta)} f(\varphi, \vartheta) \, d\Omega, \quad (\text{A.22})$$

(Driscoll and Healy Jr., 1994, p.203) Figure A.2 illustrates a complete set of the spherical harmonic functions for the levels $n \in \{0, 1, 2, 3\}$. The spherical harmonic addition theorem is defined as follows. If the angles satisfy the identity:

$$\cos(\theta) = \cos \varphi \cos \varphi_0 \sin \varphi \sin \varphi_0 \cos(\vartheta - \vartheta_0), \quad (\text{A.23})$$

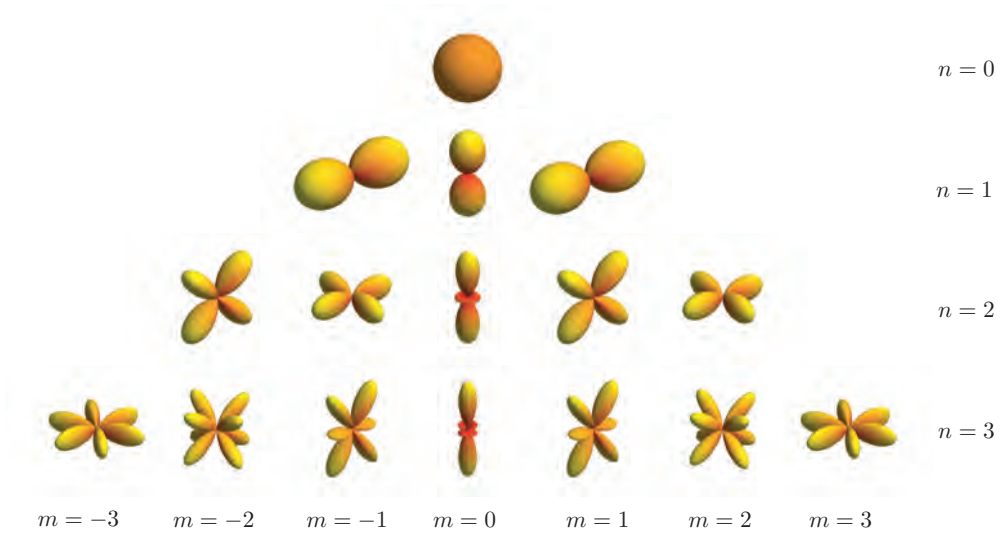


Figure A.2: Complete set of spherical harmonics $Y_n^m(\vartheta, \varphi)$ for the levels $n \in \{0, 1, 2, 3\}$. The squared real parts $\text{Re}\{Y_n^m(\vartheta, \varphi)\}^2$ are plotted (Thiergart, 2007).

the additions theorem states that (Arfken and Weber, 2005)

$$\frac{2n+1}{4\pi} P_n(\cos \theta) = \sum_{m=-n}^n Y_n^m(\varphi, \vartheta) \overline{Y_n^m(\varphi, \vartheta)}. \quad (\text{A.24})$$

A.4 Air Absorption

The numerical values of absorption due to air as used in the simulation within this thesis are calculated using the following expression (Boone and de Vries, 2004)

$$\alpha_{air}(f, s) = 8.686 f^2 \left[1.84 \cdot 10^{-11} \frac{p_{0ref}}{p_0} \sqrt{\frac{T}{R_{ref}}} + \left(\frac{T}{R_{ref}} \right)^{-\frac{5}{2}} \cdot \left(\frac{0.01275 e^{-\frac{3352}{T}}}{f_{r0} + f^2/f_{r0}} + \frac{0.1068 e^{-\frac{3352}{T}}}{f_{rN} + f^2/f_{rN}} \right) \right], \quad (\text{A.25})$$

with $p_{0ref} = 1013.25$ hPa and $T_{ref} = 293.15$ K. The temperature is denoted by T and f represents the frequency. For practical situations $T = T_0$ and $p_{0ref} = p_0$ is assumed. f_{r0} and f_{rN} are defined as

$$f_{r0} = \frac{p_0}{p_{0ref}} \left(24 + 4.04 \cdot 10^4 \frac{h(0.02 + h)}{0.391 + h} \right) \quad (\text{A.26})$$

and

$$f_{rN} = \frac{p_0}{p_{0ref}} \sqrt{\frac{T_{ref}}{T}} \left(9 + 280he^{-4.170 \sqrt[3]{\frac{T_{ref}}{T-1}}} \right). \quad (\text{A.27})$$

In Eq. (A.27), h denotes the molar concentration of water vapor in %. Because in practical situations the relative humidity h_r is measured, the following expression is used:

$$h = \frac{p_{0ref}}{p_0} \cdot h_r \cdot 10^{8.422 - 10.06 \frac{T_{ref}}{T} - 5.023 \log \frac{T}{T_{ref}} + 23 \cdot 10^{-4.44 \frac{T_{ref}}{T}}} \quad (\text{A.28})$$

The calculation gives the frequency-dependent air absorption $\alpha_{air}(f, s)$ in $\frac{dB}{m}$.

A.5 Coherence of Arbitrary Stereophonic Microphone Setups

Coherence Measurements An important property of a specific stereophonic microphone setup is the inter-channel coherence characteristic. The correlation and coherence measurements are calculated based on the derivations in Section 2.7. Now the spatial correlation and coherence in reverberant acoustic fields are extended to microphones with arbitrary first-order directivities for the 2D case. Based on the results of Kuster (2008) the more general 3D case is derived here. It is used to characterize different stereo main microphone setups and compare the results of virtual microphone setups to real setups. For this purpose 3D formulations of the spatio-temporal coherence function and correlation function are derived. The signal of two first-order microphones $S_1(\mathbf{x}_1, t)$ and $S_2(\mathbf{x}_2, t)$ positioned at the points P_1 and P_2 described by the vectors \mathbf{x}_1 and \mathbf{x}_2 can be written using equation (5.13):

$$S_1(\mathbf{x}_1, t) = \beta_{p_1} p(\mathbf{x}_1, t) + \beta_{v_1} \rho_0 c \mathbf{v}_1(\mathbf{x}_1, t) \quad (\text{A.29a})$$

$$S_2(\mathbf{x}_2, t) = \beta_{p_2} p(\mathbf{x}_2, t) + \beta_{v_2} \rho_0 c \mathbf{v}_2(\mathbf{x}_2, t), \quad (\text{A.29b})$$

where $\beta_{p_{\langle \cdot \rangle}}$ describes the directivity of the microphone and $\beta_{v_{\langle \cdot \rangle}} = (1 - \beta_{p_{\langle \cdot \rangle}})$. The pressure and velocity for the two points are denoted by $p(\mathbf{x}_{\langle \cdot \rangle}, t)$ and $\mathbf{v}(\mathbf{x}_{\langle \cdot \rangle}, t)$, respectively. The velocity vector can be expressed in separated components for the axes of Cartesian coordinates, including a rotation of the microphone in azimuth $\varphi_{\langle \cdot \rangle}$ and elevation $\vartheta_{\langle \cdot \rangle}$ as

$$\begin{aligned} S_1(\mathbf{x}_1, t) = & \beta_{p_1} p(\mathbf{x}_1, t) + \beta_{v_1} \cos(\varphi_1) \sin(\vartheta_1) X(\mathbf{x}_1, t) \\ & + \beta_{v_1} \sin(\varphi_1) \sin(\vartheta_1) Y(\mathbf{x}_1, t) + \beta_{v_1} \cos(\vartheta_1) Z(\mathbf{x}_1, t) \end{aligned} \quad (\text{A.30a})$$

$$\begin{aligned} S_2(\mathbf{x}_2, t) = & \beta_{p_2} p(\mathbf{x}_2, t) + \beta_{v_2} \cos(\varphi_2) \sin(\vartheta_2) X(\mathbf{x}_2, t) \\ & + \beta_{v_2} \sin(\varphi_2) \sin(\vartheta_2) Y(\mathbf{x}_2, t) + \beta_{v_2} \cos(\vartheta_2) Z(\mathbf{x}_2, t). \end{aligned} \quad (\text{A.30b})$$

As shown in Section 2.7, it is necessary to decompose the two signals into terms which are proportional to the pressure and velocity components parallel and perpendicular to the vector \mathbf{r} between the two measurement positions. For this purpose the two angles φ_{12} and ϑ_{12} have to be taken into account. Using the

geometry from Figure 2.8 on page 28, the factor β_v can be decomposed as

$$\beta_{v_{\parallel},1} = \beta_{v_1} \cos(\varphi_1 - \varphi_r) \sin(\vartheta_1 - \vartheta_r) \quad (\text{A.31a})$$

$$\beta_{v_{\perp 1},1} = \beta_{v_1} \sin(\varphi_1 - \varphi_r) \sin(\vartheta_1 - \vartheta_r) \quad (\text{A.31b})$$

$$\beta_{v_{\perp 2},1} = \beta_{v_1} \cos(\vartheta_1 - \vartheta_r). \quad (\text{A.31c})$$

Now the microphone signals can be written in terms of pressure and particle velocity components parallel and perpendicular to the vector \mathbf{r} as

$$S_1(0, t) = \beta_{p1}p(0, t) + \rho_0 c \beta_{v_{\perp 1},1} v_{\perp 1}(0, t) + \rho_0 c \beta_{v_{\perp 2},1} v_{\perp 2}(0, t) + \rho_0 c \beta_{v_{\parallel},1} v_{\parallel}(0, t) \quad (\text{A.32a})$$

$$S_2(\mathbf{r}, t) = \beta_{p2}p(\mathbf{r}, t) + \rho_0 c \beta_{v_{\perp 1},2} v_{\perp 1}(\mathbf{r}, t) + \rho_0 c \beta_{v_{\perp 2},2} v_{\perp 2}(\mathbf{r}, t) + \rho_0 c \beta_{v_{\parallel},2} v_{\parallel}(\mathbf{r}, t). \quad (\text{A.32b})$$

The basis for the calculation of the spatial correlation coefficient function are the cross-correlation and auto-correlation functions $R_{12}(r, \tau)$, $R_{11}(0, 0)$ and $R_{22}(0, 0)$. The cross-correlation functions can be calculated by the means of the expected value operator. Due to its linearity it can be calculated componentwise (Kuster, 2008). For notation convenience the following terms are left out because they are zero: $R_{v_{\perp},p}(r, \tau)$, $R_{v_{\perp},v_{\parallel}}(r, \tau)$, $R_{v_{\parallel},v_{\perp}}(r, \tau)$ and $R_{p,v_{\perp}}(r, \tau)$ as derived in (Jacobson, 1979).

$$\begin{aligned} R_{12}(r, \tau) = & (\rho_0 c)^2 \cdot \beta_{v_{\perp 1},1} \beta_{v_{\perp 1},2} \cdot R_{v_{\perp 1,1},v_{\perp 1,2}}(r, \tau) \\ & + (\rho_0 c)^2 \cdot \beta_{v_{\perp 2},2} \beta_{v_{\perp 2},1} \cdot R_{v_{\perp 2,2},v_{\perp 2,1}}(r, \tau) \\ & + (\rho_0 c)^2 \cdot \beta_{v_{\parallel},2} \beta_{v_{\parallel},1} \cdot R_{v_{\parallel 2},v_{\parallel 1}}(r, \tau) \\ & + \rho_0 c \cdot \beta_{v_{\parallel},2} \beta_{p,1} \cdot R_{v_{\parallel 2},p_1}(r, \tau) \\ & + \rho_0 c \cdot \beta_{v_{\parallel},1} \beta_{p,2} \cdot R_{p_2,v_{\parallel 1}}(r, \tau) \\ & + \beta_{p,1} \beta_{p,2} \cdot R_{p_2,p_1}(r, \tau) \end{aligned} \quad (\text{A.33})$$

To calculate the correlation coefficient function $\rho_{12}(r, \tau)$, Eq. (A.33) has to be normalized by the values of the autocorrelation function at $r = 0$ and $\tau = 0$

given by

$$\begin{aligned}
 R_{11}(0, 0) &= \beta_{p1}^2 \cdot R_{p1,p1}(0, 0) & (A.34a) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\perp 1,1}}^2 \cdot R_{v_{\perp 1,1}, v_{\perp 1,1}}(0, 0) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\parallel 1}}^2 \cdot R_{v_{\parallel 1}, v_{\parallel 1}}(0, 0) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\perp 2,1}}^2 \cdot R_{v_{\perp 2,1}, v_{\perp 2,1}}(0, 0)
 \end{aligned}$$

and

$$\begin{aligned}
 R_{22}(0, 0) &= \beta_{p2}^2 \cdot R_{p2,p2}(0, 0) & (A.34b) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\perp 1,2}}^2 \cdot R_{v_{\perp 1,2}, v_{\perp 1,2}}(0, 0) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\parallel 2}}^2 \cdot R_{v_{\parallel 2}, v_{\parallel 2}}(0, 0) \\
 &+ (\rho_0 c)^2 \cdot \beta_{v_{\perp 2,2}}^2 \cdot R_{v_{\perp 2,2}, v_{\perp 2,2}}(0, 0).
 \end{aligned}$$

Using the results from Eq. (2.75e) and Eq. (2.75f), the following expressions can be found

$$R_{11}(0, 0) = G^2 \left(\frac{\beta_{p1}^2}{2} + \frac{\beta_{v_{\perp 1,1}}^2}{6} + \frac{\beta_{v_{\perp 2,1}}^2}{6} + \frac{\beta_{v_{\parallel 1}}^2}{6} \right) \quad (A.35a)$$

$$R_{22}(0, 0) = G^2 \left(\frac{\beta_{p2}^2}{2} + \frac{\beta_{v_{\perp 1,2}}^2}{6} + \frac{\beta_{v_{\perp 2,2}}^2}{6} + \frac{\beta_{v_{\parallel 2}}^2}{6} \right), \quad (A.35b)$$

where G is an arbitrary scaling parameter. Using the above equations and the correlation coefficient functions between components of the particle velocity, and/or the pressure as given in Eq. (2.74), the spatial correlation coefficient

between microphones S_1 and S_2 in a diffuse field is given by

$$\begin{aligned}
 \rho_{12}(r, \tau) &= \frac{1}{\sqrt{K_1 K_2}} & (A.36) \\
 &\cdot \left[\beta_{v_{\perp 1}, 1} \beta_{v_{\perp 1}, 2} \cdot \frac{\rho_{v_{\perp 1}, 1, v_{\perp 1}, 2}(r, \tau)}{3} \right. \\
 &+ \beta_{v_{\perp 2}, 2} \beta_{v_{\perp 2}, 1} \cdot \frac{\rho_{v_{\perp 2}, 2, v_{\perp 2}, 1}(r, \tau)}{3} \\
 &+ \beta_{v_{\parallel}, 2} \beta_{v_{\parallel}, 1} \cdot \frac{\rho_{v_{\parallel}, 2, v_{\parallel}, 1}(r, \tau)}{3} \\
 &+ \beta_{v_{\parallel}, 2} \beta_{p, 1} \cdot \frac{\rho_{v_{\parallel}, 2, p_1}(r, \tau)}{\sqrt{3}} \\
 &+ \beta_{v_{\parallel}, 1} \beta_{p, 2} \cdot \frac{\rho_{p_2, v_{\parallel}, 1}(r, \tau)}{\sqrt{3}} \\
 &\left. + \beta_{p, 1} \beta_{p, 2} \cdot \rho_{p_2, p_1}(r, \tau) \right],
 \end{aligned}$$

with

$$K_1 = \beta_{p_1}^2 + \frac{\beta_{v_{\perp 1}, 1}^2}{3} + \frac{\beta_{v_{\perp 2}, 1}^2}{3} + \frac{\beta_{v_{\parallel}, 1}^2}{3}, \quad (A.37a)$$

$$K_2 = \beta_{p_2}^2 + \frac{\beta_{v_{\perp 1}, 2}^2}{3} + \frac{\beta_{v_{\perp 2}, 2}^2}{3} + \frac{\beta_{v_{\parallel}, 2}^2}{3}. \quad (A.37b)$$

For the derivation of the spatial coherence function, (Kuster, 2008) has proposed the following approach. Using the correlation function, the spectral densities can be found utilizing Eq. (2.72). Based on the spectral density, the coherence function $\gamma^2(r, \omega)$ can be calculated. The correlation function given in Eq. (2.75) contains the time-dependent factors $\sin(\omega\tau)$ and $\cos(\omega\tau)$. The Fourier transform is given by

$$\mathcal{F}\{\sin(\omega_0, \tau)\} = i\pi [\delta(\omega + \omega_0) - \delta(\omega - \omega_0)] \quad (A.38a)$$

and

$$\mathcal{F}\{\cos(\omega_0, \tau)\} = \pi [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]. \quad (A.38b)$$

Interpreting the delta functions as defining the spectral densities at the frequencies ω_0 , the cross-spectral density $S_{12}(r, \omega)$ for positive frequencies be-

comes (Kuster, 2008)

$$\begin{aligned}
 S_{12}(r, \omega) = & \frac{\pi G^2}{2} (\beta_{v_{\perp 1}, 1} \beta_{v_{\perp 1}, 2} \cdot \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} & (A.39) \\
 & + \beta_{v_{\perp 2}, 2} \beta_{v_{\perp 2}, 1} \cdot \frac{\sin(kr) - (kr) \cos(kr)}{(kr)^3} \\
 & + \beta_{v_{\parallel}, 2} \beta_{v_{\parallel}, 1} \cdot \frac{(kr^2) \sin(kr) + 2kr \cos(kr) - 2 \sin(kr)}{(kr)^3} \\
 & + \beta_{p, 1} \beta_{p, 2} \cdot \frac{\sin(kr)}{kr}) \\
 & + i \frac{\pi G^2}{2} (\beta_{v_{\parallel}, 2} \beta_{p, 1} \cdot (\frac{\sin(kr) - (kr) \cos(kr)}{(kr)^2}) \\
 & + \beta_{v_{\parallel}, 1} \beta_{p, 2} \cdot (\frac{\sin(kr) - (kr) \cos(kr)}{(kr)^2})), & (A.40)
 \end{aligned}$$

while utilizing Eq. (2.75). The auto spectral densities are given by

$$S_{11}(0, \omega) = \frac{\pi G^2}{2} K_1 \quad (A.41)$$

and

$$S_{11}(0, \omega) = \frac{\pi G^2}{2} K_2. \quad (A.42)$$

Using Eq. (A.39) - (A.42) the spatial coherence function between the two microphones is given by

$$\gamma_{12}^2(\omega) = \frac{|S_{12}(r, \omega)|^2}{S_{11}(0, \omega) S_{22}(0, \omega)}. \quad (A.43)$$

Appendix B

Additional Plots

B.1 Plots Specific to Section 3.5.2

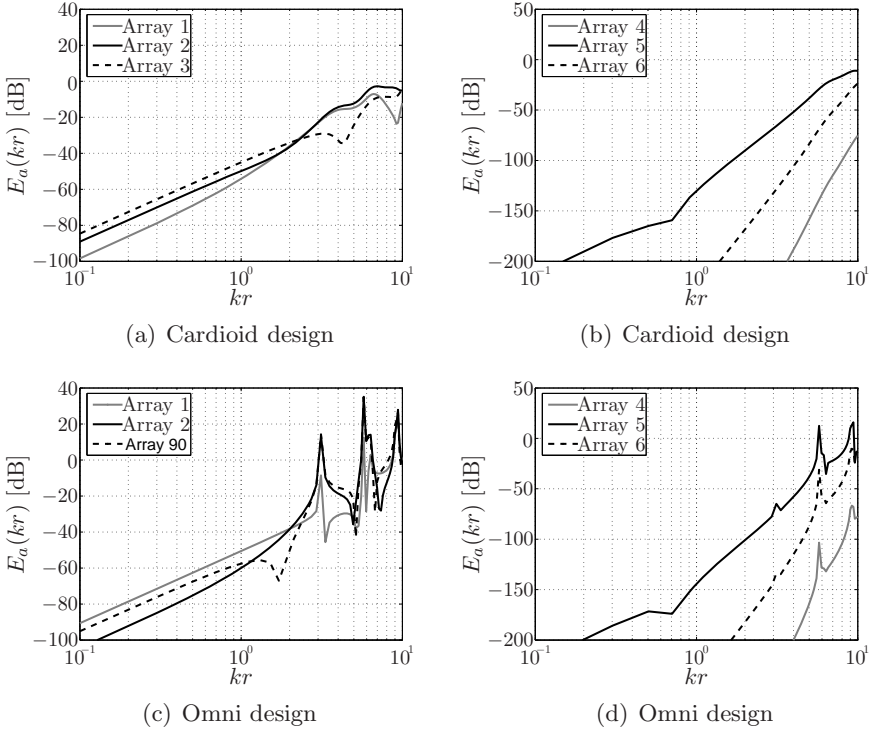
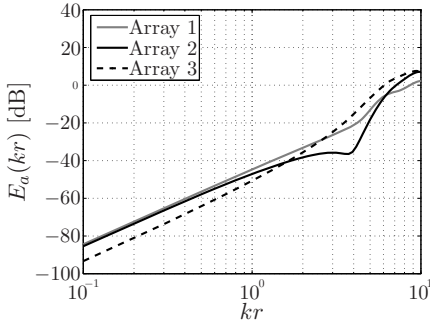
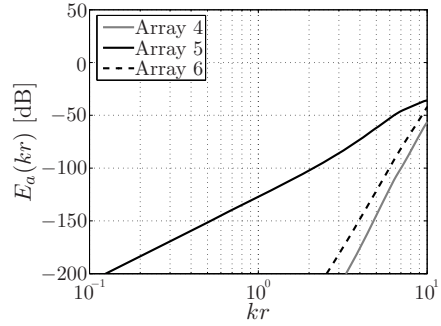


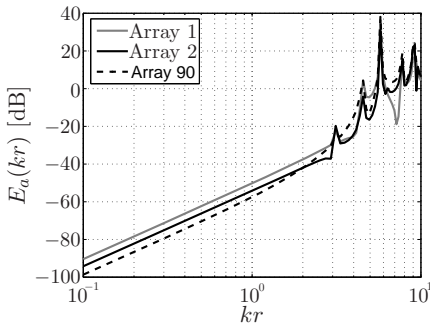
Figure B.1: Relative error contribution $10\log_{10}(E_a(kr))$ of spatial aliasing for the array configurations in Table 3.2, but for the arrival direction $\Omega_0 = (0^\circ, 90^\circ)$ Thiergart (2007).



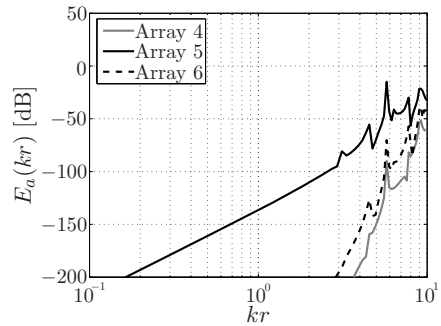
(a) Cardiod design



(b) Cardiod design



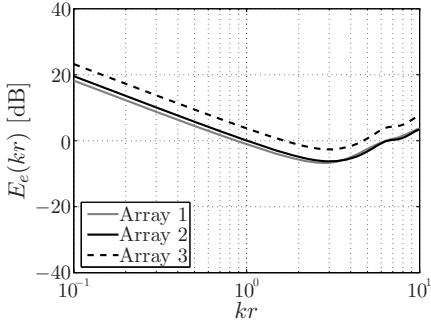
(c) Omni design



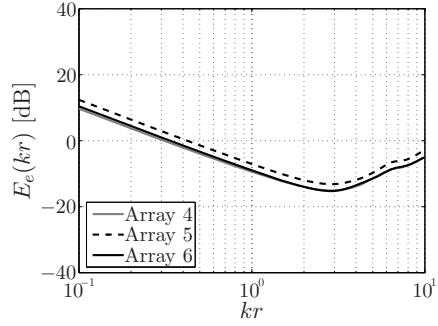
(d) Omni design

Figure B.2: Relative error contribution $10 \log_{10}(E_a(kr))$ of spatial aliasing for the array configurations in Table 3.2, but for the arrival direction $\Omega_0 = (-165^\circ, 22^\circ)$ Thiergart (2007).

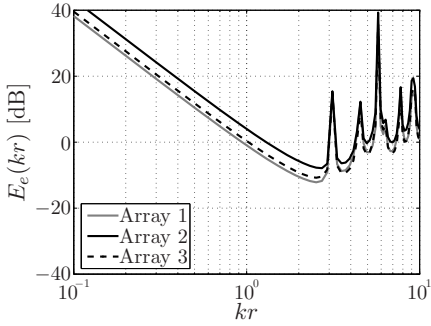
B.2 Plots Specific to Section 3.5.3



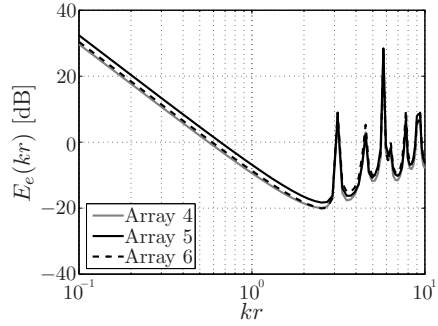
(a) Cardiod design



(b) Cardiod design

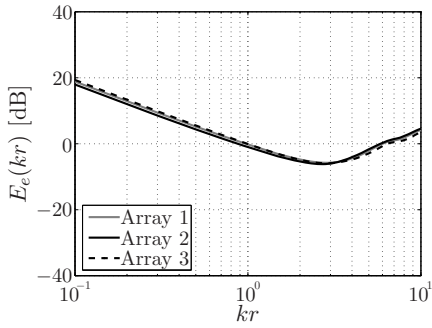


(c) Omni design

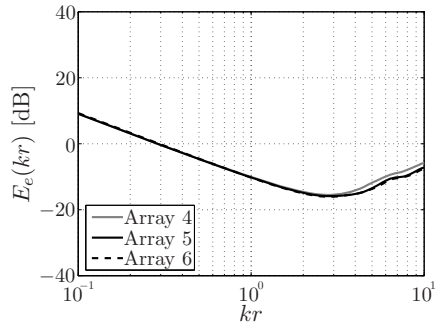


(d) Omni design

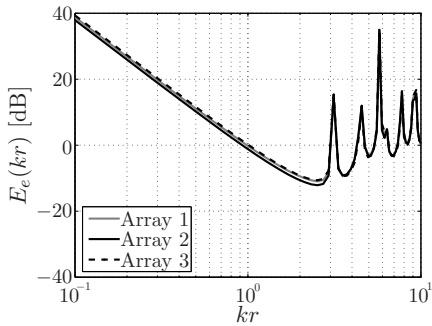
Figure B.3: Relative error contribution $10 \log_{10}(E_e(kr))$ of microphone noise for the array configurations in Tab. 3.2, but for the arrival direction $\Omega_0 = (0^\circ, 90^\circ)$ (Thiergart, 2007).



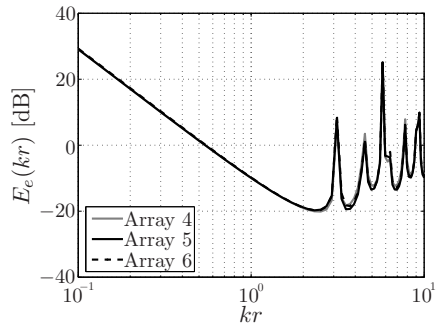
(a) Cardiod design



(b) Cardiod design



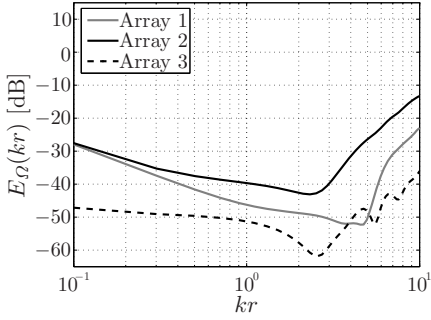
(c) Omni design



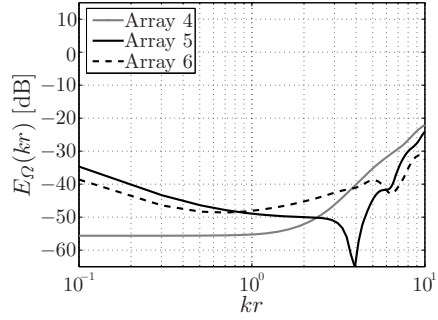
(d) Omni design

Figure B.4: Relative error contribution $10 \log_{10}(E_e(kr))$ of microphone noise for the array configurations in Tab.3.2, but for the arrival direction $\Omega_0 = (-165^\circ, 22^\circ)$ (Thiergart, 2007).

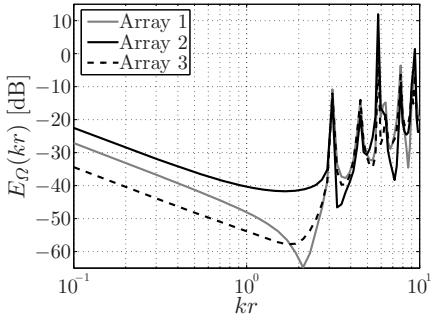
B.3 Plots Specific to Section 3.5.4



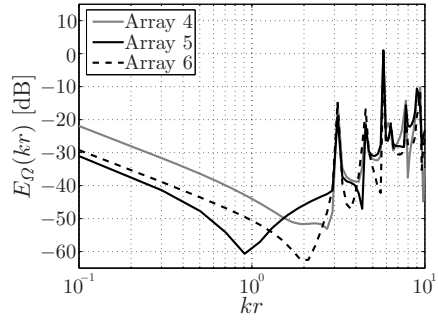
(a) Cardiod design



(b) Cardiod design

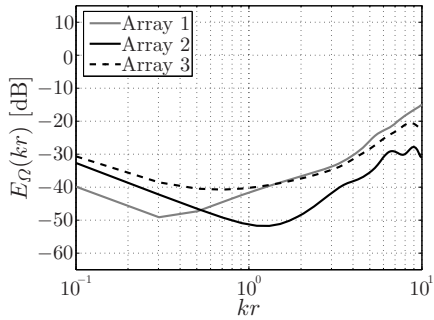


(c) Omni design

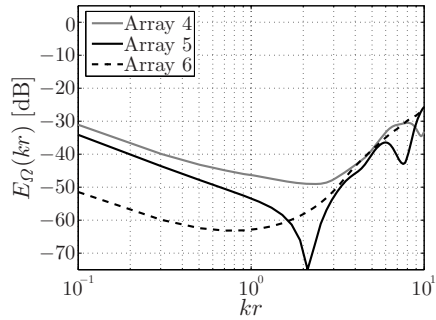


(d) Omni design

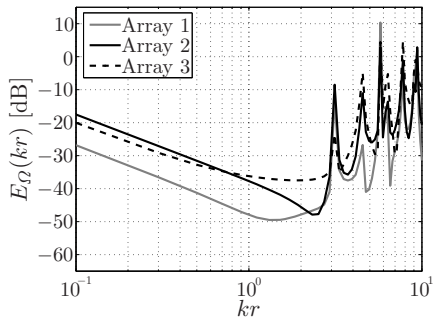
Figure B.5: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of random positioning errors for the array configurations in Tab. 3.2, but for the arrival direction $\Omega_0 = (0^\circ, 90^\circ)$ (Thiergart, 2007).



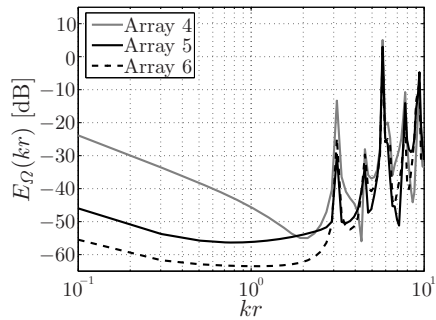
(a) Cardiod design



(b) Cardiod design

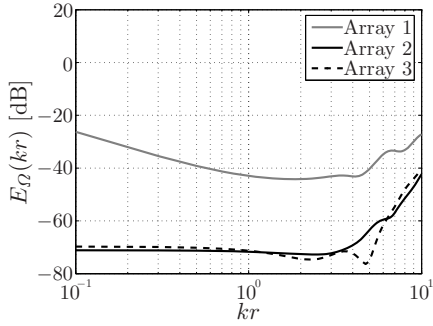


(c) Omni design

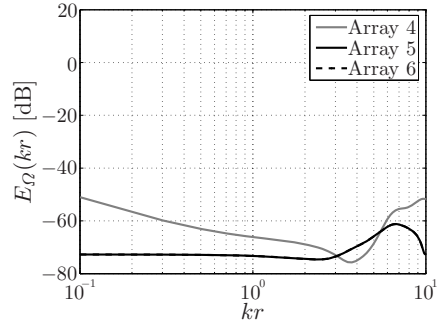


(d) Omni design

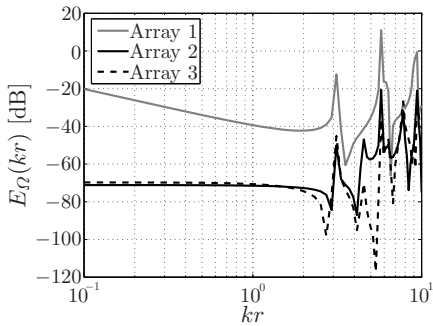
Figure B.6: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of random positioning errors for the array configurations in Tab. 3.2, but for the arrival direction $\Omega_0 = (-165^\circ, 22^\circ)$ (Thiergart, 2007).



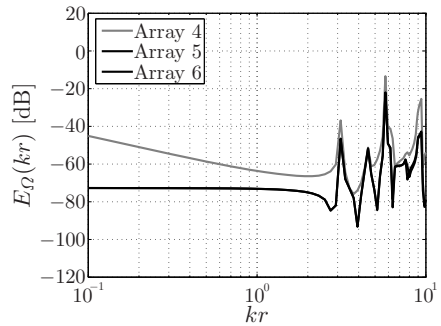
(a) Cardiod design



(b) Cardiod design

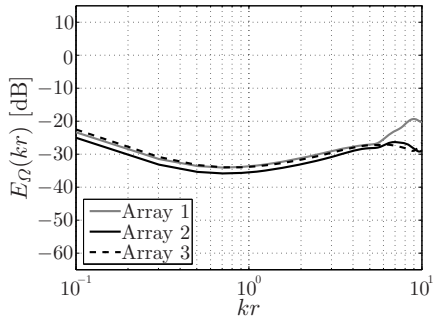


(c) Omni design

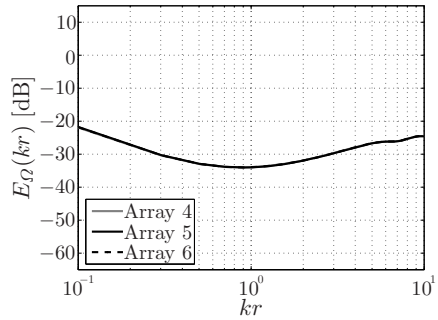


(d) Omni design

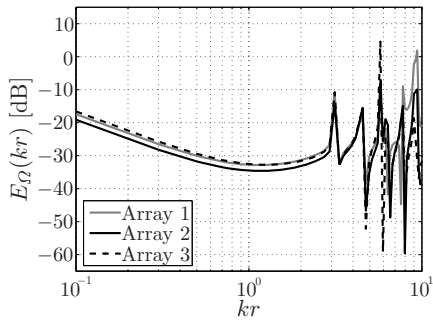
Figure B.7: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of vertical offset positioning errors for the array configurations in Tab.3.2, but for the arrival direction $\Omega_0 = (0^\circ, 90^\circ)$ (Thiergart, 2007).



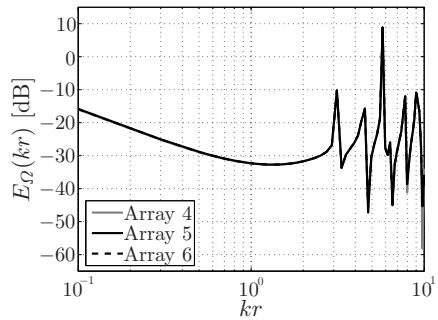
(a) Cardiod design



(b) Cardiod design



(c) Omni design



(d) Omni design

Figure B.8: Relative error contribution $10 \log_{10}(E_{\Omega}(kr))$ of vertical offset positioning errors for the array configurations in Tab. 3.2, but for $\Omega_0 = (-165^\circ, 22^\circ)$ (Thiergart, 2007).

B.4 Plots Specific to Section 3.5.5

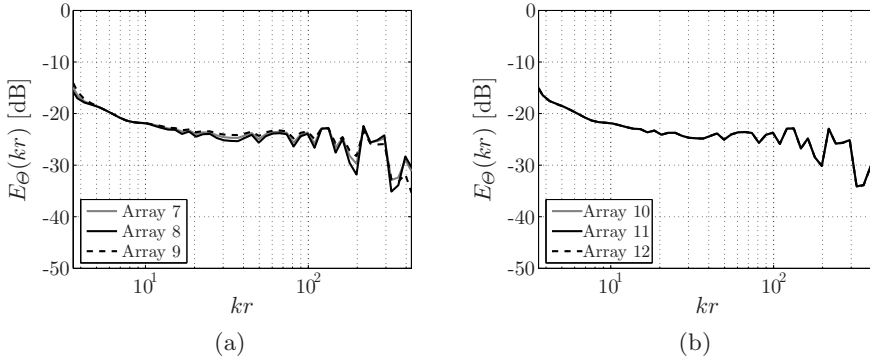


Figure B.9: Relative error contribution $10 \log_{10} E_{\Theta}(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a constant radius $r = 1.5$ m for $\Omega_0 = (0^\circ, 90^\circ)$ (Thiergart, 2007).

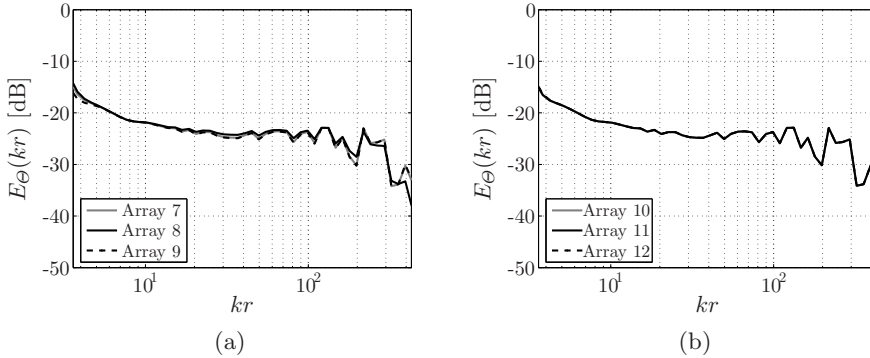


Figure B.10: Relative error contribution $10 \log_{10} E_{\Theta}(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a radius $r = 1.5$ m for $\Omega_0 = (-165^\circ, 22^\circ)$ (Thiergart, 2007).

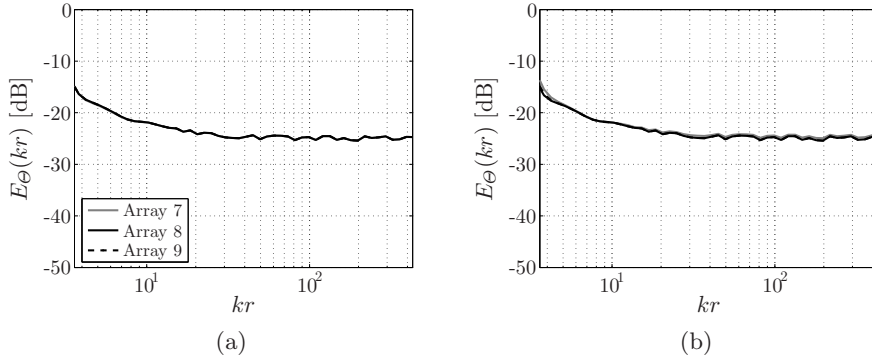


Figure B.11: Relative error contribution $10 \log_{10} E_{\Theta}(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a constant frequency $f = 1$ kHz $\Omega_0 = (0^\circ, 90^\circ)$ (Thiergart, 2007).

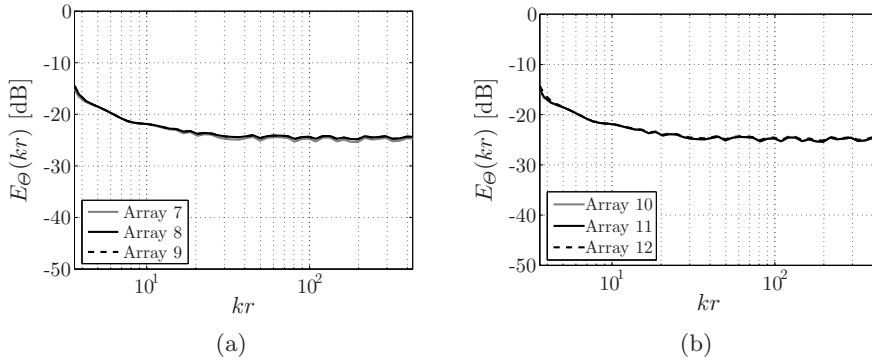


Figure B.12: Relative error contribution $10 \log_{10} E_{\Theta}(kr)$ of a non-ideal cardioid microphone characteristic (MKH 8040) for a frequency $f = 1$ kHz $\Omega_0 = (-165^\circ, 22^\circ)$ (Thiergart, 2007).

Appendix C

Glossary of Acronyms

2D	two-dimensional
3D	three-dimensional
AAE	area of accurate extrapolation
AB	stereo microphone setup
AR	augmented reality
BL	Blumlein stereo microphone setup
DFT	discrete Fourier transform
DSTFT	discrete short time Fourier transform
DIX	directivity index
DI	dynamic interaction
DOA	direction of arrival
DOF	degree of freedom
EDC	energy decay curve
EDR	energy decay relief
EQ	equalizer
FT	Fourier transform
FS	Fourier series
FFT	fast Fourier transform
FIR	finite impulse response
GUI	graphical user interface
HRTF	head-related transfer function
HRIR	head-related impulse response
HMD	head-mounted display
ISTFT	inverse short time Fourier transform
ISFT	inverse spherical Fourier transform

IDSTFT	discrete short time Fourier transform
IR	impulse response
IDFT	inverse discrete Fourier transform
IFT	inverse Fourier transform
IACC	inter-aural cross-correlation coefficient
ITD	inter-aural time difference
ILD	inter-aural level difference
IIR	infinite impulse response
LTI	linear time-invariant
MISM	mirror image source model
MUSHRA	multiple stimuli with hidden reference and anchor
NURBS	Non-Uniform Rational B-Splines
NSR	noise-to-signal ratio
ORTF	stereo microphone setup named after the office de radio-diffusion télévision française
PWD	plane wave decomposition
RIR	room impulse response
RMSE	root mean squared error
SFT	spherical Fourier transform
STFT	short time Fourier transform
SI	static interaction
SNR	signal-to-noise ratio
SHD	spherical harmonic decomposition
TOA	time of arrival
TF	transfer function
VR	virtual reality
VM	virtual microphones
WFS	wave field synthesis
WFE	wave field extrapolation
WNG	white-noise-gain
XML	extensible markup language

Bibliography

- Abel, J. and Huang, P. (2006). A simple, robust measure of reverberation echo density. Convention paper presented at the 121st AES Convention, San Francisco.
- Abhayapala, T. D. and Ward, D. (2002). Theory and design of high order sound field microphones using spherical microphone array. *IEEE Transactions on acoustics, speech and signal processing*.
- Ahrens, J. (2010). *The Single-layer Potential Approach to Sound Field Synthesis Including Cases of Non-enclosing Distributions of Secondary Sources*. PhD thesis, Technische Universität Berlin.
- Ahrens, J. and Spors, S. (2007). Implementation of directional sources in wave field synthesis. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 66–69.
- Albrecht, B. (2005). Personal communication.
- Alexander, R. C. (2000). *The Life and Works of Alan Dower Blumlein*. Focal Press.
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). The CIPIC HRTF database. In *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pages 99–102.
- Allen, J. B. (1977). Short term spectral analysis, synthesis and modification by discrete transform. *IEEE Transactions on acoustics, speech and signal processing*, ASSP-3(3):235–238.
- Allen, J. B. and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 66(4):943–950.
- Allen, J. B. and Rabiner, L. R. (1977). A unified approach to short-time Fourier

- analysis and synthesis. In *Proceedings of the IEEE*, volume 65, pages 1558–1564.
- Altiverb (2009). *Altiverb Convolution Reverb*. Internet resource, available at <http://www.audioease.com/Pages/Altiverb/AltiverbMain.html> (accessed: 28th January 2009).
- Arfib, D., Keiler, F., and Zölzer, U. (2007). *DAFX Digital Audio Effects*, chapter Time-frequency Processing. John Wiley & Sons.
- Arfken, G. B. and Weber, H. J. (2005). *Mathematical Methods for Physicists*. Elsevier Academic Press.
- Avni, A. and Rafaely, B. (2009). Interaural cross correlation and spatial correlation in a sound field represented by spherical harmonics. Conference paper presented at the Ambisonics Symposium Graz.
- Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6, 4(6):355–385.
- Baalman, M. A. J. (2008). *On wave field synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily shaped sound sources*. PhD thesis, Technische Universität Berlin.
- Balakrishnan, R. and Kurtenbach, G. (1999). Exploring bimanual camera control and object manipulation in 3D graphic interfaces. In *Proceedings of 1999 ACM Conference on Human Factors in Computing System (CHI'99)*, pages 56–63.
- Balmages, I. and Rafaely, B. (2007). Open sphere design for spherical microphone arrays. *IEEE Transactions on audio, speech and language processing*, 15(2).
- Bäni, W. (2002). *Wavelets - Eine Einführung für Ingenieure*. Oldenbourg.
- Bartlett, B. (1991). *Stereo microphone techniques*. Focal Press.
- Begault, D. R. (1992). Perceptual effects of synthetic reverberation on 3-D audio systems. *J. Audio Eng. Soc.*, 40(11):895–903.
- Begault, D. R., Wenzel, E. M., and Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916.

- Bendat, J. and Piersol, A. (1980). *Engineering Applications of Correlation and Spectral Analysis*. Wiley.
- Berkhout, A. J. (1988). A holographic approach to acoustic control. *J. Audio Eng. Soc.*, 36(12):977–995.
- Berkhout, A. J. and de Vries, D. (1989). Acoustic holography for sound control. Convention paper presented at the 86th AES Convention, Hamburg.
- Berkhout, A. J., de Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764 – 2778.
- Bitzer, J. and Simmer, K. U. (2001). *Microphone Arrays*. Springer.
- Blackstock, D. T. (2000). *Fundamentals of Physical Acoustics*. Wiley & Sons.
- Blauert, J. (1997). *Spatial Hearing*. The MIT Press.
- Blumlein, A. D. (1931). Improvements in and relating to sound-transmission, sound-recording and sound-reproduction systems. In *B.P. 394 325*.
- Boone, M. M., de Bruijn, W. P. J., and Horbach, U. (1999). Virtual surround speakers with wave field synthesis. Convention paper presented at the 106th AES Convention, Munich.
- Boone, M. M. and de Vries, D. (2004). Lecture notes: Sound Control Part 2 Ap 3551 TU Delft.
- Boor, C. D. (1994). *A Practical Guide to Splines*. Springer.
- Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America*, 75(6):1828–1836.
- Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. (2005). *3D User Interfaces - Theory and Practice*. Addison-Wesley.
- Braasch, J. (2005). A loudspeaker-based 3D sound projection using virtual microphone control (vimic). Convention paper presented at the 118th AES Convention, Barcelona.
- Brandstein, M. and Ward, D. (2001). *Microphone Arrays*. Springer.
- Bronstein, I., Semendjajew, K., G.Musiol, and Mühlig, H. (1995). *Taschenbuch der Mathematik*. Verlag Harri Deutsch.

- Chen, J., Veen, B. D. V., and Hecox, K. E. (1992). External ear transfer function modeling: A beamforming approach. *Journal of the Acoustical Society of America*, 92(4):1933–1944.
- Chun, I., Rafaely, B., and Joseph, P. (2003). Experimental investigation of spatial correlation in broadband reverberant sound fields. *Journal of the Acoustical Society of America*, 113(4):1995–1998.
- Collins, R. E. (1998). *Mathematical Methods for Physicists and Engineers*. Dover Publications, 2nd edition.
- Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., and JR., M. C. T. (1955). Measurement of correlation coefficients in reverberant sound fields. *Journal of the Acoustical Society of America*, 27(6):1072–1077.
- Corteel, E. (2007). Synthesis of directional sources using wave field synthesis, possibilities, and limitations. *EURASIP Journal on Advances in Signal Processing*, 2007.
- Cremer, L. and Müller, H. A. (1976). *Die wissenschaftlichen Grundlagen der Raumakustik*. Hirzel.
- Crochiere, R. (1980). A weighted overlap-add method of short-time fourier analysis/synthesis. *IEEE Transactions on acoustics, speech and signal processing*, ASSP-28(1):99–102.
- Daniel, J., Nicol, R., and Moreau, S. (2003). Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. Convention paper presented at the 114th AES Convention, Paris.
- de Bruijn, W. P. J. (2004). *Application of Wave Field Synthesis in videoconferencing*. PhD thesis, Delft University of Technology.
- de Vries, D., Baan, J., and Sonke, J.-J. (1997). Multi-channel wave field analysis as a specification tool for wave field synthesis. Convention paper presented at the 103rd AES Convention, Munich.
- de Vries, D., Hulsebos, E., and Baan, J. (2000). Spatial fluctuation of spaciousness measures in auditoria. Convention paper presented at the 108th AES Convention, Paris.
- Dickreiter, M. (2008). *Handbuch der Tonstudioteknik Band 1*. K G Sauer.
- Driscoll, J. R. and Healy Jr., D. M. (1994). Computing Fourier transforms and

- convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15(2):202–250.
- Duda, R. O. and Martens, W. L. (1998). Range dependence or the response of a spherical head model. *Journal of the Acoustical Society of America*, 104(5):3048–3058.
- Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S. (2005). High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues. Convention paper presented at the 119th AES Convention, New York.
- Elko, G. (2001). *Microphone Arrays*. Springer.
- Engeln-Müllges, G., Niederdrenk, K., and Wodicka, R. (2005). *Numerik- Algorithmen - Verfahren, Beispiele, Anwendungen*. Springer Berlin.
- Falch, C., Noisternig, M., Warum, S., and Hoeldrich, R. (2003). Room simulation for binaural sound reproduction using measured spatialtemporal impulse responses. In *6th Int. Conference on Digital Audio Effects (DAFX-03)*.
- Farina, A. and Ayalon, R. (2003). Recording concert hall acoustics for posterity. Conference paper presented at the 24th AES Conference, Banff.
- Farina, A. and Trochnin, L. (2004). Advanced techniques for measuring and reproducing spatial sound properties of auditoria. In *International Symposium on Room Acoustics Design and Science*.
- Fisher, E. and Rafaely, B. (2008). The nearfield spherical microphone array. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5272–5275.
- Forsyth, M. (1985). *Buildings for music: The architect, the musician and the listener from the 17th century to the present day*. MIT Press.
- Freitag, K. . (2000). *Kling & Freitag CA 106*. Product Sheet, available at http://www.kling-freitag.info/downloads/dokumente/008_Datenbl%C3%A4tter/dat_CA106.pdf (accessed: 5th February 2008).
- Frohlich, D. (1993). The history and future of direct manipulation. Technical report, HP Laboratories Bristol.
- Galdo, G. D. (2007). *Geometry-based Channel Modeling for Multi-User MIMO Systems and Applications*. PhD thesis, Technische Universität Ilmenau.

- Gerzon, M. A. (1973). Periphony: With-height sound reproduction. *Journal of the AES*, 21(1):2–10.
- Giron, F. (1996). *Investigations about the directivity of sound sources*. PhD thesis, Ruhr-Universität Bochum.
- Görne, T. (1996). *Mikrofone in Theorie und Praxis*. Elektor-Verlag GmbH.
- Gover, B. N., Ryan, J. G., and Stinson, M. R. (2002). Microphone array measurement system for analysis of directional and spatial variations of sound fields. *Journal Acoustic Society of America*, 112(5):1980–1991.
- Gover, B. N., Ryan, J. G., and Stinson, M. R. (2003). Designing a spherical microphone array for the directional analysis of reflections and reverberation. Convention paper presented at the 115th AES Convention, New York.
- Gover, B. N., Ryan, J. G., and Stinson, M. R. (2004). Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array. *Journal of the Acoustical Society of America*, 116(5):2138–2148.
- Gradshteyn, S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Elsevier Academic Press.
- Gramann, S. (2006). Benutzerinteraktion für räumliche Tonwiedergabesysteme auf Basis parametrisierter Impulsantworten und deren interaktiven Visualisierung. Diplomarbeit, Technische Universität Ilmenau, Fraunhofer IDMT.
- Griesinger, D. (1989). Practical processors and programs for digital reverberation. Conference paper presented at the 7th AES Conference, Toronto.
- Grossman, T. and Balakrishnan, R. (2006). The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of UIST 2006 - the ACM Symposium on User Interface Software and Technology*, pages 3 – 12.
- Guiard, Y. (1987). Symmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behaviour*, 19(4):486–517.
- Guillaume, M. and Grenier, Y. (2006a). Experimental 3D sound field analysis with a microphone array. Conference paper presented at the 28th AES Conference, Pitea.
- Guillaume, M. and Grenier, Y. (2006b). Sound field analysis with a two-

- dimensional microphone array. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.
- Hamasaki, K., Hiyama, K., and Okumura, R. (2005). The 22.2 multichannel sound system and its application. Convention paper presented at the 118th AES Convention, Barcelona.
- Hamasaki, K., Komiya, S., Okubo, H., Hiyama, K., and Hatano, W. (2004). 5.1 and 22.2 multichannel sound productions using an integrated surround sound panning system. Convention paper presented at the 117th AES Convention, San Francisco.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83.
- Helleman, H. (2003). Sensitivity of the human auditory system to spatial variations in single early reflections. Master’s thesis, Delft University of Technology.
- Herder, J. (1998). Tools and widgets for spatial sound authoring. In *Computer Network and ISDN System 30*, pages 1933–1940. Elsevier.
- Herder, J. and Novotny, T. (2003). Spatial sound design and interaction for virtual environments in the promotion of architectural designs. Proceedings of the 3rd International Workshop on Spatial Media.
- Houde, S. (1992). Interactive design of an interface for easy 3-D direct manipulation. In *ACM Conference on Human Factors in Computing Systems (CHI’02)*. ACM.
- Hulsebos, E. (2004). *Auralization using wave field synthesis*. PhD thesis, Delft University of Technology.
- Huopaniemi, J. and III, J. O. S. (1999). Spectral and time-domain preprocessing and the choice of modeling error criterion for binaural digital filters. Convention paper presented at the AES 16th International Conference, Arktikum, Rovaniemi.
- ISO 3382 (2000). Acoustics - measurement of the reverberation time of rooms with reference to other acoustical parameters. ISO Norm 3382.
- ITU (2003). ITU-R BS.1534-1 Method for the subjective assessment of intermediate quality level of coding systems. Technical report, International Telecommunication Union (ITU) Radiocommunication Assembly.

- Jacobson, F. (1979). The diffuse sound field. Technical Report 27, The Acoustic Laboratory, Technical University of Denmark.
- Jot, J.-M. (1992). An analysis/synthesis approach to real-time artificial reverberation. In *IEEE Int. Conf. Acou. Speech and Signal Proc.*, volume II, pages 221–224.
- Jot, J.-M., Cerveau, L., and Warusfel, O. (1997). Analysis and synthesis of room reverberation based on a statistical time-frequency model. Convention paper presented at the 103rd AES Convention, New York.
- Jot, J.-M., Larcher, V., and Warusfel, O. (1995). Digital signal processing issues in the context of binaural and transaural stereophony. Convention paper presented at the 98th AES Convention, Paris.
- Kahrs, M. and Brandenburg, K. (1998). *Applications of digital signal processing to audio and acoustics*. Kluwer.
- Kessler, R. (2005). An optimised method for capturing multidimensional acoustic fingerprints. Convention paper presented at the 118th AES Convention, Barcelona.
- Kircher, A. (1983 reprint from 1684). *Neue Hall- und Thonkunst*. Edition librari Th. Schaefer.
- Kistler, D. J. and Wightman, F. L. (1992). A model of head-related transfer function based on principal components analysis and minimum-phase reconstruction. *Journal Acoustic Society of America*, 91(3):1537–1647.
- Koenig, W., Dunn, H. K., and Lacy, L. Y. (1946). The sound spectrograph. *Journal Acoustic Society of America*, 18(1):19–49.
- Kosten, C. W. (1960). The mean free path in room acoustics. *Acustica*, 10:245 – 250.
- Kress, D. and Irmer, R. (1989). *Angewandte Systemtheorie*. VEB Verlag Technik Berlin.
- Kuntz, A. (2009). *Wave Field Analysis Using Virtual Circular Microphone Arrays*. PhD thesis, Universität Erlangen Nürnberg.
- Kuntz, A. and Rabenstein, R. (2008). On the extrapolation of room impulse responses from circular measurements for data based wave field synthesis. Conference paper presented at the 34th Annual German Conference on Acoustics (DAGA), Dresden.

- Kuster, M. (2008). Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity. *Journal of the Acoustical Society of America*, 123(1):154–162.
- Kuttruff, H. (1991). *Room acoustics*. Elsevier Applied Science New York.
- Laborie, A., Bruno, R., and Montoya, S. (2003). A new comprehensive approach of surround sound recording. Convention paper presented at the 114th AES Convention, Paris.
- Larcher, V., Vandernoot, G., and Jot, J.-M. (1998). Equalization methods in binaural technology. Convention paper presented at the 105th AES Convention, San Francisco.
- Laubach, T. (2004). Prototypische Entwicklung eines audiovisuellen Augmented-Reality System mit Wellenfeldsynthese. Diplomarbeit, Technische Universität Ilmenau, Fraunhofer IDMT.
- Lebedev, V. (1975). Values of the nodes and weights of ninth to seventeenth order gauss-markov quadrature formulae invariant under the octahedron group with inversion. *Computational Mathematics and Mathematical Physics*, 15:44–51.
- Lebedev, V. (1976). Quadratures on a sphere. *Computational Mathematics and Mathematical Physics*, 16(2):293–306.
- Lebedev, V. (1977). Spherical quadrature formulas exact to orders 25-29. *Siberian Mathematical Journal*, 18:99–107.
- Lebedev, V. (1995). A quadrature formula for the sphere of 59th algebraic order of accuracy. *Russian Acad. Sci. Dokl. Math.*, 50:283–286.
- Lebedev, V. and Laikov, D. (1999). A quadrature formula for the sphere of the 131st algebraic order of accuracy. *Doklady Mathematics*, 59(3):477–481.
- Lebedev, V. and Skorokhodov, A. (1992). Quadrature formulas of orders 41, 47, and 53 for the sphere. *Russian Acad. Sci. Dokl. Math.*, 45:587–592.
- Lebedev, V. I. (2009). *Fortran code for Lebedev grids*. Internet resource, available at <http://www.ccl.net/cca/software/SOURCES/FORTRAN/Lebedev-Laikov-Grids/> (accessed: 19th January 2009).
- Li, Z. and Duraiswami, R. (2005). Hemispherical microphone array for sound

- capture and beamforming. In *IEEE Workshop on Applications of Signal Processing to Audio and Multimedia*, pages 106–109.
- Li, Z. and Duraiswami, R. (2007). Flexible and optimal design of spherical microphone arrays for beamforming. *IEEE Transactions on audio, speech and language processing*, 15(2):702–714.
- Lindau, A., Kosanke, L., and Weinzierl, S. (2010). Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. Convention paper presented at the 128th AES Convention, London.
- Lokki, T. and Karjalainen, M. (2000). An auditory motivated analysis method for room impulse responses. In *COST G-6 Conference on Digital Audio Effects (DAFX-00)*.
- Mackensen, P., Feldhoff, U., Theile, G., Horbach, U., and Pellegrini, R. (1999). Binaural room scanning - a new tool for acoustic and psychoacoustic research. In *Conference paper presented at the 25th Annual German Conference on Acoustics (DAGA), Munich*.
- Meesawat, K. and Hammershoi, D. (2003). The time when the reverberation tail in a binaural room impulse response begins. Convention paper presented at the 115th AES Convention, New York.
- Melchior, F., Fischer, J.-O., and de Vries, D. (2006a). Audiovisual perception using wave field synthesis in combination with augmented reality systems: Horizontal positioning. Conference paper presented at the 28th AES Conference, Pitea.
- Melchior, F., Gatzsche, G., Reichelt, K., Deguara, J., Dausel, M., and Strauss, M. (2006b). Universal system for spatial sound reinforcement in theatres and large venues - system design and user interface. Convention paper presented at the 120th AES Convention, Paris.
- Melchior, F., Laubach, T., and de Vries, D. (2005). Authoring and user interaction for the production of wave field synthesis content in an augmented reality system. In *The fourth IEEE and ACM International Symposium on Mixed and Augmented Reality*.
- Melchior, F., Röder, T., Wabnik, S., Brix, S., and Riegel, C. (2003). Authoring systems for wave field synthesis content production. Convention paper presented at the 115th AES Convention, New York.

- Melchior, F., Sladeczek, C., Parzsch, A., and Brix, S. (2010). Design and implementation of an interactive room simulation for wave field synthesis. Presented at the AES 40th International Conference, Tokyo.
- Meltzer, S., Altmann, L., Gräfe, A., and Fischer, J.-O. (2008). An object oriented mixing approach for the design of spatial audio scenes. Convention paper presented at the 25. VDT Convention, Leipzig.
- Menzel, D., Wittek, H., Theile, G., and Fastl, H. (2005). The binaural sky: A virtual headphone for binaural room synthesis. Conference paper presented at the 1st VDT-Symposium, Germany.
- Merimaa, J. (2006). *Analysis, synthesis and perception of spatial sound - binaural localization modeling and multichannel loudspeaker reproduction*. PhD thesis, Helsinki University of Technology.
- Merimaa, J., Lokki, T., Peltonen, T., and Karjalainen, M. (2001). Measurement, analysis and visualisation of directional room responses. Convention paper presented at the 111th AES Convetion, New York.
- Meyer, J. (1995). *Akustik und musikalische Aufführungspraxis*. Bochinsky.
- Meyer, J. and Agnello, T. (2003). Sphercial microphone array for spatial sound recording. Convention paper presented at the 115th AES Convention, New York.
- Meyer, J. and Elko, G. (2002). A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *IEEE International Conference on Acoustic Speech and Signal Processing*, pages 1181–1184.
- Meyer, J. and Elko, G. (2006). Position independent close talking microphone. *Signal Processing*, 86(6):1254–1259.
- Minnaar, P., Christensen, F., Moller, H., Olesen, S. K., and Plogsties, J. (1999). Audibility of all-pass components in binaural synthesis. Convention paper presented at the 106th AES Convention, Munich.
- Minnaar, P., Olesen, S. K., Christensen, F., and Moller, H. (2001). The importance of head movements for binaural room synthesis. In *International Conference on Auditory Display (ICAD)*.
- Minnaar, P., Plogsties, J., and Christensen, F. (2005). Directional resolution

- of head-related transfer functions required in binaural synthesis. *Journal of the AES*, 53(10):919–929.
- Minnaar, P., Plogsties, J., Olesen, S. K., Christensen, F., and Moller, H. (2000). The interaural time difference in binaural synthesis. Convention paper presented at the 108th AES Convention, Paris.
- Mommertz, E. (1996). *Untersuchung akustischer Wandeigenschaften und Modellierung der Schallrueckwuerfe in der binauralen Raumsimulation*. PhD thesis, RWTH Aachen.
- Moore, B. C. J. (1995). *Hearing Handbook of perception and cognition*. Academic Press.
- Moreau, S., Daniel, J., and Bertel, S. (2006). 3D sound field recording with higher order ambisonics - objective measurements and validation of a 4th order spherical microphone. Convention paper presented at the 120th AES Convention, Paris.
- Morse, P. M. and Ingard, K. U. (1968). *Theoretical Acoustics*. Princeton University Press.
- Möser, M. (2007). *Technische Akustik*. Springer.
- Müller, S. and Massarani, P. (2001). Transfer-function measurement with sweeps. *Journal of the AES*, 49(6):443–471.
- Muraoka, T. (2006). Evaluation of ambience microphone arrangements utilizing frequency dependent spatial cross correlation (FSCC). Convention paper presented at the 120th AES Convention, Paris.
- Nicol, R. (1999). *Restitution sonore spatialisé sur une zone étendue: Application a la telepreséence*. PhD thesis, Université du Maine.
- Noll, P. (1997). MPEG Digital audio coding. *IEEE Signal Processing Magazine*, pages 59–81.
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999). *Discrete-Time Signal Processing*. Prentice Hall Signal Processing Series.
- Paris, I. (2008). *The LISTEN HRTF database*. IRCAM Paris. Internet resource, available at <http://recherche.ircam.fr/equipes/salles/listen/index.html> (accessed: 15th October 2008).

- Park, M. and Rafaely, B. (2005). Sound-field analysis by plane-wave decomposition using spherical microphone array. *Journal of the Acoustical Society of America*, 118(5):3094–3103.
- Pellegrini, R. (2001). *A Virtual Reference listening Room as an Application of Auditory Virtual Enviroments*. PhD thesis, Ruhr-Universität Bochum.
- Piegl, L. (1991). On nurbs: A survey. *IEEE Computer Graphics and Applications*, 11(1):55 – 71.
- Piegl, L. and Tiller, W. (1997). *The NURBS Book*. Springer.
- Pierce, A. D. (1981). *Acoustics: An introduction to its physical principles and applications*. Mc Graw-Hill.
- Poletti, M. A. (2005). Three-dimensional surround sound systems based on spherical harmonics. *J. Audio Eng. Soc.*, 53(11):1004–1024.
- Potard, G. (2006). *3D-Audio Object Oriented Coding*. PhD thesis, University of Wollongong.
- Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466.
- Rafaely, B. (2000). Spatial-temporal correlation of a diffuse sound field. *Journal Acoustic Society of America*, 107(6):3254–3258.
- Rafaely, B. (2004). Plane-wave decomposition of the sound field on a sphere by spherical convolution. *Journal of the Acoustical Society of America*, 116(4).
- Rafaely, B. (2005a). Analysis and design of spherical microphone arrays. *IEEE Transactions on speech and audio processing*, 13(1):135–143.
- Rafaely, B. (2005b). Phase-mode versus delay-and-sum spherical microphone array processing. *IEEE Signal Processing Letters*, 12(10).
- Rafaely, B. (2008a). Spatial sampling and beamforming for spherical microphone arrays. In *Hands-Free Speech Communication and Microphone Arrays HSCMA*, pages 5–8.
- Rafaely, B. (2008b). The spherical-shell microphone array. *IEEE Transactions on audio, speech and language processing*, 16(4):740–747.
- Rafaely, B., Balmages, I., and Eger, L. (2007a). High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical mi-

- crophone array. *Journal of the Acoustical Society of America*, 122(5):2661–2668.
- Rafaely, B., Weiss, B., and Bachmat, E. (2007b). Spatial aliasing in spherical microphone arrays. *IEEE Transaction on signal processing*, 55(3):1003–1010.
- Raskin, J. (2000). *The Humane Interface*. Addison-Wesley.
- Reverb, C. (2009). *IR1 Parametric Convolution Reverb*. Waves. Internet resource, available at www.waves.com/content.aspx?id=250 (accessed: 29th January 2009).
- Rieländer, M. M. (1982). *Reallexikon der Akustik*. Verlag Erwin Bachinsky.
- Roper, S. and Collins, T. (2007). A sound source and reflections localization method for reverberant rooms using arrays of microphones. Conference paper presented at the 32nd AES Conference, Hillerod.
- Rubak, P. and Johansen, L. (1998). Artificial reverberation based on a pseudo-random impulse response I. Convention paper presented at the 104th AES Convention.
- Rubak, P. and Johansen, L. (1999). Artificial reverberation based on a pseudo-random impulse response II. Convention paper presented at the 106th AES Convention.
- Rumsey, F. (2001). *Spatial audio*. Focal Press.
- Sabine, W. C. (1964). *Collected papers on acoustics*. Peninsula.
- Sandvad, J. and Hammershoi, D. (1994). Binaural auralization. Comparison of FIR and IIR filter representation of HIRs. Convention paper presented at the 96th AES Convention, Amsterdam.
- Scheck, A. (2008). Entwicklung eines Interaktionsmodells zur Manipulation der Richtcharakteristik einer Schallquelle in VR- und Desktop-Benutzerschnittstellen. Diplomarbeit, Technische Universität Ilmenau, Fraunhofer IDMT.
- Schlesinger, A. (2006). Arraytechnologie in der Raumakustik. Diplomarbeit, Technische Universität Ilmenau, Fraunhofer IDMT.
- Schlesinger, A., Albrecht, B., Galdo, G. D., Husung, S., and Lotze, J. (2007). Holographic sound field analysis with a scalable spherical microphone-array. Convention paper presented at the 122nd AES Convention, Vienna.

- Schmalstieg, D., Fuhrmann, A., and Hesina, G. (2002). The Studierstube augmented reality project. *PRESENCE - Teleoperators and Virtual Environments*, 11(1).
- Schneidmadel, F. and Seidenzahl, K. (2005). Untersuchungen zu neuen User Interfaces und Paradigmen für die objektorientierte Produktion von auditiven Szenen. Medienprojekt TU Ilmenau, Fraunhofer IDMT.
- Schroeder, M. (1962). Natural sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3).
- Schroeder, M. (1965). New method for measuring reverberation time. *Journal of the Acoustical Society of America*, 37:409–412.
- Sennheiser (2007a). *Sennheiser MKH 20 P 48*. Product Sheet, available at [http://www.sennheiser.com/sennheiser/icm_eng.nsf/resources/MKH_20_P_48_GB.pdf/\\$File/MKH_20_P_48_GB.pdf](http://www.sennheiser.com/sennheiser/icm_eng.nsf/resources/MKH_20_P_48_GB.pdf/$File/MKH_20_P_48_GB.pdf) (accessed: 28th October 2007).
- Sennheiser (2007b). *Sennheiser MKH 30 P 48*. Product Sheet, available at [http://www.sennheiser.com/sennheiser/icm_eng.nsf/resources/MKH_30_P_48_GB.pdf/\\$File/MKH_30_P_48_GB.pdf](http://www.sennheiser.com/sennheiser/icm_eng.nsf/resources/MKH_30_P_48_GB.pdf/$File/MKH_30_P_48_GB.pdf) (accessed: 28th October 2007).
- Sennheiser (2007c). *Sennheiser MKH 8040*. Internet resource, available at <http://www.sennheiser.com/sennheiser/icm.nsf/root/500966#> (accessed: 28th October 2007).
- Serra, X. (1989). *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic decomposition*. PhD thesis, CCRMA Stanford University.
- SGI (2008). *Open Inventor*. Internet resource, available at <http://oss.sgi.com/projects/inventor/> (accessed: 12th October 2008).
- Silzle, A., Novo, P., and Strauss, H. (2004). Ika-sim: A system to generate auditory virtual environments. Convention paper presented at the 116th AES Convention, Berlin.
- Sladeczek, C. (2008). Untersuchungen zur nutzerspezifischen Optimierung der Wellenfeldsynthese. Diplomarbeit, Bauhaus-Universität Weimar, Fraunhofer IDMT.

- Snow, W. B. (1953). Basic principles of stereophonic sound. *Journal of the SMPTE*, 61:567–587.
- Sonke, J.-J. (2000). *Variable acoustics by wave field synthesis*. PhD thesis, Delft University of Technology.
- Spors, S. (2005). *Active Listening Room compensation for Spatial Sound Reproduction Systems*. PhD thesis, Universität Erlangen Nürnberg.
- Spors, S. (2007). Extension of an analytic secondary source selection criterion for wave field synthesis. Convention paper presented at the 123rd AES Convention, New York.
- Springer, J. P., Sladeczek, C., Scheffler, M., Hochstrate, J., Melchior, F., and Froehlich, B. (2006). Combining wave field synthesis and multi-viewer stereo displays. In *Proceedings of the IEEE Virtual Reality Conference (VR'06)*.
- Stanley, M. (1993). An efficient implementation of the Patterson-Holdsworth auditory filter bank. Perception Group - Advanced Technology Group 35, Apple Computer.
- Start, E. (1997). *Direct Sound Enhancement by Wave Field Synthesis*. PhD thesis, Delft University of Technology.
- Steinberg, J. and Snow, W. B. (1934). Auditory perspective - physical factors. *Electrical Engineering*, pages 12 – 17.
- Stoakley, R., Cornway, M., and Pausch, R. (1995). Virtual reality on a wim: interactive worlds in miniature. In *SIGCHI Conference on Human Factors in Computing Systems*.
- Sutherland, I. E. (2003). Sketchpad: A man-machine graphical communication system. Technical report, University of Cambridge.
- Teutsch, H. (2007). *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Springer.
- Theile, G. (2001). Multichannel natural music recording based on psychoacoustic principles. Technical report, Institut für Rundfunktechnik.
- Thiele, R. (1953). Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen. *Acustica*, page 291.
- Thiergart, O. (2007). Sound field analysis on the basis of a spherical micro-

- phone array for auralization applications. Diplomarbeit, Technische Universität Ilmenau, Fraunhofer IDMT.
- van Dorp Schuitman, J. and de Vries, D. (2008). Applying cochlear modeling and psychoacoustics in room acoustics. Convention paper presented at the 124th AES Convention, Amsterdam.
- van Trees, H. L. (2002). *Optimum array processing*. Wiley-Interscience.
- Varshalovich, D., Moskalev, A., and Khersonskii, V. (1988). *Quantum Theory of Angular Momentum*. World Scientific.
- Veen, B. D. V. and Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *ASSP Magazin*, 5(2):4–24.
- Verheijen, E. N. G. (1998). *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology.
- Vogel, P. (1993). *Application of wave field synthesis in room acoustics*. PhD thesis, Delft University of Technology.
- Vorländer, M. (2008). *Auralization*. Springer.
- Ware, C. (2004). *Information visualization*. Elsevier.
- Williams, E. (1999). *Fourier Acoustics*. Academic Press.
- Williams, M. and Du, G. L. (2000). Multichannel microphone array design. Convention paper presented at the 108th AES Convention.
- Witteck, H. (2002). OPSI a hybrid WFS / phantom source solution to avoid spatial aliasing. Technical report, Institut für Rundfunktechnik.
- Wuttke, J. (2008). Mikrofonaufsätze (<http://www.schoeps.de>). Technical report, Schoeps.
- Yewdall, D. L. (1999). *Practical art of motion picture sound*. Focal Press.
- Ziomek, L. J. (1995). *Fundamentals of acoustic field theory and space-time signal processing*. CRC Press.

Summary

Developments in the area of spatial sound reproduction have led to a large variety of established audio systems. Systems based on stereophonic principles are extended and growing from two channels via the ITU-R BS.775 surround setup to larger systems with more channels including elevated loudspeaker. On the other hand, sound field reproduction systems aiming to reconstruct an acoustic field like Wave Field Synthesis (WFS) and Ambisonics, are on the verge of being available on the market. Additionally, binaural reproduction is established especially for simulation and auralization applications and psychoacoustic research and is now entering the mass market by the success of smart phones and other devices. All these systems are termed as spatial audio reproduction systems.

Using spatial reproduction systems only very few applications are aiming for a natural reproduction of a recorded situation. In most cases the aim is to communicate artistic messages or ideas:

- A recorded music performance transformed to a spatial reproduction by sound engineers.
- A pure virtual piece of music (e.g., pop music produced in a studio or electronic music).
- A virtual piece of acoustic art (e.g., radio drama).
- An audio-visual artwork (e.g., a movie and its corresponding sound track).

Most applications do not reproduce a real acoustic environment. The spatial audio scene is a pure virtual construct. The development and realization of such a scene is termed sound design. The sound designer tries to communicate an acoustical idea and needs to transform his abstract concept into acoustic reality in a given environment with a given reproduction system. Such a concept of sound is not necessarily described by the use of physical models in terms of geometrical room models with an arrangement of real sound sources. Furthermore such an acoustic idea does not and should not depend on a particular

reproduction system.

During the last decades of audio signal processing development, a countless number of tools for the modification of single audio streams have been developed (e.g., equalizer, compressor, modulation effects like chorus). All these tools can be used to modify a property of an single audio stream. The sound designer transforms his acoustic idea into a parameter set for processing devices to reach his goals of acoustic communication. Besides the artistic knowledge a strong background in signal processing and the interaction of both is required. Especially, the perceptual effect of a modification of a property of an audio stream is the key element in the know-how of a sound engineer, sound designer or Tonmeister.

In addition, the process of spatial sound design modifies the spatial properties of an audio stream including its position, direction, orientation in a virtual room and the acoustic characteristics or the room itself. For the sound design process two possible models are:

1. The virtual acoustic scene is referred to as an object oriented virtual reality composed of sound objects and sound manipulating objects (e.g. walls). An acoustic environment as it can be found in the physical world is modeled.
2. The virtual acoustic scene and corresponding acoustic field are visualized (direction dependent) in terms of direction dependent perceptual or physical properties but without a representation of physical possible objects or sound sources.

The first model limits the sound designer by physical constraints, which are part of the scene description and have to be implemented on basis of simulations. Moreover the sound designer has to adapt his acoustic idea to simulated objects. A more intuitive way to modify the sound field is a direct interaction with a graphical representation as given in the second approach.

The main objective of this thesis is to develop a spatial sound design system, which is not bounded by descriptions of physical and geometrical room acoustics and of which the interaction principles are reproduction system independent. The focus of this work lies on the sound design in terms of acoustic environments, but the introduced principles can easily be extended. It is important to notice that the basic material is the analysis of existing room or sound field. Using the principles of spatial sound design these can be modified. At this point a geometrical or physical correspondence is not required.

In this work a novel processing chain for spatial sound design based on measured room impulse responses has been developed. The system is based on spherical array measurements of room impulse responses. New interaction methods for sound designers have been developed based on such measurements. The interaction principles developed are independent of specific reproduction systems and based on direct interaction with visualizations of spatial impulse responses. Following this processing chain each building block has been analyzed in detail.

In the acquisition of spatial impulse responses the properties of cardioid open sphere virtual microphone arrays were investigated in terms of error robustness and spatial sampling. Virtual means in this context that the impulse responses are measured consecutively with a single microphone on a robotic arm. The capabilities of extrapolation and plane wave decomposition of such a system were analyzed. The combination of simultaneous measurements on different radii to extend the usable frequency range was investigated and implemented.

The analysis of measured room impulse responses has been described and a storage format for the analyzed spatial impulse response has been developed and used in the realized processing framework. The extraction of single events (e.g., reflections) from spherical array measurements was studied by measurements in an anechoic room equipped with a single reflecting surface. It was shown that with adequate spatio-temporal filtering the frequency response of a reflection can be extracted.

A taxonomy of impulse response visualization was given in order to develop a reproduction system independent interaction method for spatial sound design. Suitable techniques for the direct interaction with different visualized direction dependent impulse responses were investigated and implemented. The interaction process was classified in static and dynamic interactions. The static interaction of the spatial sound design process modifies the parts of the room impression which are independent from a source position and a specific reproduction system. The dynamic interaction process is depending on the source and reproduction system configuration.

For the static interaction new methods like inverse energy decay curve editing and the concept of shaping surfaces were developed and applied. Time-variant filtering, based on the short-time Fourier transform and the spatial envelope shaping are new principles studied in this context. The dynamic interaction was analyzed and methods for an efficient design of spatial acoustic scenes were developed. The interaction techniques that are proposed have been real-

ized within a graphical user interface for desktop use. An extension of a new user interface has been described and prototypes have been realized using an augmented reality framework and state-of-the-art user interface hardware.

The dynamic interaction was studied for wave field synthesis, stereophonic reproduction and binaural reproduction. Methods for the auralization and adaptation of measured and modified high resolution data to these three reproduction systems were developed. The effects of measurement errors on the reproduction quality were investigated using listening experiments employing simulated measurement data based on mirror image source models and diffuse field simulations. The advantages of dual radius cardioid spherical microphone arrays and the developed adaptation methods for stereo, binaural, and WFS reproduction have been demonstrated.

The methods proposed and developed in this thesis are based on the measurement of room impulse responses. The next step is the application of the proposed methods to array recordings of complete performances. To make this possible a multichannel array recording and analysis system with very high spatial resolution has to be developed. Since the complexity and costs are very high for real time recording compared with the room impulse response measurements used in this thesis a model which optimizes such a system based on the human perception will be very important.

The author believes that a tempo-spatial editing of recorded events, will be available in the future and extend the freedom of the sound design process in a unprecedented way. One can think of a simultaneous recording of a acoustic event where the spatial arrangement and the properties of the different sources can be changed and edited in the post production process without recording each source separately.

Samenvatting

Ontwikkelingen op het gebied van ruimtelijke geluidweergave hebben geleid tot een grote variëteit van gevestigde audiosystemen. Op stereofonische principes gebaseerde systemen zijn uitgebreid van twee kanalen, via de ITU-R BS.775 surround setup, tot grote systemen met vele kanalen, inclusief luidsprekers boven het horizontale vlak. Anderzijds staan weergavesystemen die de reconstructie van een geluidveld beogen - zoals golfveldsynthese (wave field synthesis, WFS) en ambisonics - op het punt om door te breken op de markt. Daarnaast is er de binaurale reproductie, speciaal ontwikkeld voor toepassingen op het gebied van simulatie, auralisatie en psycho-akoestische research. Deze vindt zijn weg op de consumentenmarkt door het succes van smartphones en andere apparatuur. Deze systemen worden ruimtelijke audio-reproductiesystemen genoemd.

Slechts heel weinig toepassingen van ruimtelijke reproductiesystemen beogen een geregistreerde akoestische situatie natuurgetrouw weer te geven. Meestal is het doel om artistieke boodschappen of ideeën door te geven, zoals:

- een opname van een muziekkuitvoering die door geluidstechnici tot ruimtelijke weergave is omgewerkt;
- een zuiver virtueel stuk muziek (b.v. popmuziek geproduceerd in een studio, of elektronische muziek);
- een virtueel akoestisch kunstwerk (b.v. radiodrama);
- een audio-visueel kunstwerk (b.v. een film en de bijbehorende soundtrack).

De meeste toepassingen reproduceren geen werkelijke akoestische omgeving, maar een zuiver virtuele 'audio scene'. De ontwikkeling en realisatie van zo'n scene wordt geluidontwerp genoemd. De geluidontwerper probeert een akoestisch idee weer te geven en moet zijn abstracte concept omvormen tot een akoestische werkelijkheid, in een gegeven omgeving en met een gegeven weergavesysteem. Zijn geluidconcept wordt niet noodzakelijkerwijs beschreven door

geometrische ruimtemodellen en een configuratie van werkelijke geluidbronnen. Bovendien hangt zijn geluidconcept niet af van een bepaald weergavesysteem. Dat moet ook niet.

In de laatste decennia zijn talloze 'tools' ontwikkeld voor de modificatie van individuele 'audio streams' zoals de equalizer, de compressor en modulatie-effecten zoals chorus. Deze tools kunnen worden gebruikt om een bepaalde eigenschap van een enkele audio stream te veranderen. Om de gewenste akoestische communicatie tot stand te brengen vertaalt de geluidontwerper zijn akoestische idee in een set parameters om zijn tools aan te sturen. Naast artistieke kundigheid is hierbij een gedegen kennis van signaalverwerking vereist en ook moet de interactie tussen beide disciplines worden begrepen. Het belangrijkste element in de know-how van een geluidtechnicus, geluidontwerper of 'Tonmeester' is het inzicht in de perceptieve effecten die een modificatie van een eigenschap van een audio stream veroorzaakt.

Het proces van ruimtelijk geluidontwerp verandert de ruimtelijke eigenschappen van een audio stream, zoals zijn positie, richting en oriëntatie in een virtuele ruimte, en de eigenschappen van die ruimte zelf. Twee modellen zijn mogelijk voor ruimtelijk geluidontwerp:

1. De virtuele acoustic scene wordt beschouwd als een object-georiënteerde 'virtual reality', samengesteld uit geluidobjecten en objecten die het geluid manipuleren (b.v. muren). Zo wordt een akoestische omgeving gemodelleerd zoals in werkelijkheid kan worden aangetroffen.
2. De virtuele acoustic scene en het daarmee corresponderende geluidveld worden richtingafhankelijk gevisualiseerd, in termen van richtingafhankelijke perceptieve of fysische eigenschappen, maar zonder een representatie van fysische mogelijke objecten of geluidbronnen.

Het eerste model beperkt de geluidontwerper door fysische voorschriften die onderdeel zijn van de beschrijving van de audio scene en die op basis van simulaties moeten worden geïmplementeerd. Daarnaast moet de geluidontwerper zijn akoestische idee aan gesimuleerde objecten aanpassen. Een meer intuïtieve manier om het geluidveld te veranderen is directe interactie met een grafische representatie ervan, zoals geboden door het tweede model.

Het voornaamste doel van dit proefschrift is de ontwikkeling van een ruimtelijk geluidontwerpsysteem dat niet wordt beperkt door beschrijvingen van fysische en geometrische akoestische ruimten en waarvan de interactieprincipes onafhankelijk zijn van het weergavesysteem. De focus van het werk is gericht op

geluidontwerp in termen van akoestische omgevingen, maar de voorgestelde principes kunnen gemakkelijk worden uitgebreid. Het is belangrijk te bedenken dat het basismateriaal bestaat uit de analyse van een geluidveld in een bestaande ruimte. Gebruikmakend van de principes van ruimtelijk geluidontwerp kan dit materiaal worden veranderd tot voorbij het punt waar geometrische en fysische overeenkomsten vereist zijn.

De in dit proefschrift beschreven nieuwe bewerkingsketen voor ruimtelijk geluidontwerp is gebaseerd op pulsresponsies van een ruimte, gemeten met sferische (d.i. bolvormige) microfoonarrays. Deze metingen vormen de basis voor nieuwe interactiemethoden voor geluidontwerpers, onafhankelijk van een specifieke weergavesysteem. Ze gaan uit van directe interacties met visualisaties van de ruimtelijke pulsresponsies. Elke stap van de bewerkingsketen is in detail geanalyseerd.

Bij de acquisitie van ruimtelijke pulsresponsies zijn de eigenschappen van virtuele (d.w.z. met één of enkele microfoons serieel samengestelde) open sferische microfoonarrays onderzocht in termen van foutbestendigheid en ruimtelijke bemonstering. De mogelijkheden van extrapolatie en vlakke golf-ontbinding van deze systemen zijn onderzocht. De combinatie van simultane metingen met arrays van verschillende diameter, met als doel het bruikbare frequentiegebied uit te breiden, is onderzocht en toegepast. De analyse van gemeten pulsresponsies is beschreven. Een geschikt opslagformaat voor het analyseresultaat is ontwikkeld en toegepast in het bewerkingsproces. De selectie van individuele elementen, b.v. een enkele reflectie, uit metingen met een sferisch microfoonarray is onderzocht aan de hand van metingen in een dode kamer met daarin een enkel reflecterend oppervlak. Aangetoond is dat met adequate filtering in het ruimte-tijd domein de frequentieresponsie van zo'n reflectie kan worden bepaald.

Een taxonomie van de visualisatie van pulsresponsies is gegeven met als doel een interactiemethode voor ruimtelijk geluidontwerp te ontwikkelen die onafhankelijk is van het reproductiesysteem. Geschikte technieken werden ontwikkeld voor directe interactie met gevisualiseerde, richtingafhankelijke pulsresponsies. Het interactieproces wordt onderscheiden in statische en dynamische interactie. De statische interactie van het ruimtelijk geluidontwerpproces verandert die delen van de ruimte-indruk die onafhankelijk zijn van de bronpositie en van een specifiek reproductiesysteem. Het dynamische interactieproces hangt af van de configuraties van de bronnen en het reproductiesysteem.

Voor de statische interactie zijn nieuwe methoden ontwikkeld en toegepast,

zoals inversie van de energie-afnamecurve en het concept van 'shaping surfaces'. Tijdvariante filtering, gebaseerd op short-time fouriertransformatie en shaping van de ruimtelijke omhullende zijn nieuwe principes die in dit verband zijn onderzocht. De dynamische interactie is geanalyseerd en methoden voor een efficiënt ontwerp van ruimtelijke akoestische omgevingen werden ontwikkeld. De voorgestelde interactietechnieken zijn gerealiseerd met een grafisch 'user desktop interface'. Een nieuw, uitgebreid user interface is beschreven en prototypes zijn gerealiseerd met gebruik van een 'augmented reality' - omgeving op een geavanceerde computer.

De dynamische interactie werd onderzocht voor golfveldsynthese, stereofonische weergave en binaurale reproductie. Methoden voor auralisatie van gemeten en gemodificeerde hoge-resolutie data, aangepast aan deze drie weergavesystemen, zijn ontwikkeld. De effecten van meetfouten op de weergavekwaliteit zijn onderzocht aan de hand van luisterproeven waarbij gesimuleerde data werden gebruikt, gebaseerd op spiegelbronmodellen en diffuus-veld simulaties. De voordelen van dubbel-diameter sferische cardioide-microfoonarrays, en van de ontwikkelde methoden voor aanpassing aan de diverse weergavesystemen zijn gedemonstreerd.

De methoden die in dit proefschrift zijn voorgesteld en ontwikkeld zijn gebaseerd op het meten van pulsresponsies van ruimten. De volgende stap is toepassing van de voorgestelde methoden op 'real-time' arrayopnamen van complete 'performances'. Om dit mogelijk te maken dient een meerkanaals arrayopname- en analysesysteem met zeer hoge ruimtelijke resolutie te worden ontwikkeld. Gezien de hoge complexiteit en kosten van zo'n systeem in vergelijking met de pulsresponsiemetingen die in dit proefschrift zijn gebruikt zal de ontwikkeling van een model dat dit systeem optimaliseert met inachtneming van de menselijke perceptie zeer belangrijk zijn.

De auteur gelooft dat bewerking in het ruimte-tijd domein van opgenomen producties in de toekomst mogelijk zal zijn, waarmee de vrijheid van het geluidontwerpproces op een tot dusver ongekende manier wordt vergroot. Men kan hierbij denken aan een simultane opname van een akoestische performance waarvan in nabewerking de ruimtelijke configuratie en de eigenschappen van de verschillende bronnen kunnen worden veranderd zonder dat die bronnen afzonderlijk zijn opgenomen.

Curriculum Vitae

Name: Frank Melchior
Birth: 30th of May, 1975, in Wesel, Germany
Nationality: German

Education

2003 Degree: Dipl.-Ing. Media Technology
(equivalent to M.Sc. Media Technology)
1996 - 2003 Technical University of Ilmenau, Germany
Department of Media Technology
1995 - 1996 Technical University of Ilmenau, Germany
Department of Electrical Engineering
1994 - 1995 Alternative Service
1994 Degree: Abitur
1991 - 1994 Höhere Berufsfachschule Wesel, Germany

Profession

2009 - Director Research and Development
IOSONO GmbH, Erfurt
2003 - 2009 Researcher, Project Manager
Fraunhofer Institute Digital Media Technology
2004 - 2009 Ph.D. Candidate
Delft University of Technology, The Netherlands

