Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording

vorgelegt von
Benjamin Bernschütz, M.Sc.
aus Wesel

von der Fakultät I – Geistes- und Bildungswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Ulf Schrader
Berichter:	Prof. Dr. Stefan Weinzierl
Berichter:	Prof. Dr. Christoph Pörschmann
Berichter:	Prof. Dr. Sascha Spors

Tag der wissenschaftlichen Aussprache: 15.02.2016

Berlin 2016

Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording

Benjamin Bernschütz

Doctoral Dissertation

Research and writing from 2009–2015, Published in 2016

Technical University of Berlin – Audio Communication Group
Technology Arts Sciences TH Köln – Institute of Communications Systems

This dissertation was written at Technical University of Berlin, Faculty I – Audio Communication Group under supervision of Prof. Dr. Stefan Weinzierl.

Funded by the German Federal Ministry of Education and Research, support codes: 1707X08 and 17009X11, granted to TH Köln, Faculty 07 – Institute of Communications Systems, Prof. Dr. Pörschmann.

Written with a LaTeX typesetting system using TeX Live and Texmaker.
References managed with Papers for MAC and BibTeX.
Vector drawings made with TikZ and OmniGraffle.
Pixel graphics/photos edited with Pixelmator.
Diagrams plotted with MATLAB®.

Cover picture taken by Philipp Stade at WDR Köln.

To my beloved parents

Monika Bernschütz-Hörnchen† and Walter Bernschütz†.

Acknowledgements

I thank my doctoral adviser Professor Dr. Stefan Weinzierl for his fantastic support, fruitful discussions and invaluable advises. I am impressed by the breadth of his knowledge and his extraordinary sense of accuracy. It is a great honor and privilege to be your doctoral student.

I thank Professor Dr. Christoph Pörschmann, an incredible visionary and razor-sharp mind, who made this work possible and had the initial ideas. I had the great chance and honor to work with him for six years in an outstanding working environment. Professor Pörschmann has supported this work and myself in so many different ways. I thank you very much, on a professional as well as on a personal level.

I thank Professor Dr. Sascha Spors, one of the leading experts in spatial audio signal theory, for his steady support and guidance since the time I was writing my master thesis under his supervision. It is a great honor to have you as supporter and reviewer for my doctoral thesis now.

This work was funded by the Bundesministerium für Bildung und Forschung (BMBF) in Germany in the scope of two different research projects. Thanks for the great support and the opportunities.

Many thanks go to (in alphabetic order) Johannes Arend, Antje Goldenberg, Gary Grutzek, Dominic Hemmer, Ulla Jantzen, Philipp Stade and Arnau Vázquez-Giner, my dear colleagues and friends at the Institute of Communications Systems at TH Köln for the incredibly great times we had and for the steady professional exchange, support and inspiration.

Very special thanks for direct contributions to Arnau Vázquez-Giner who collaborated with me in the research projects. He was responsible for implementing the listening experiment software and setting up the listening experiments that appear in this work. Very special thanks to Philipp Stade and Max Rühl who performed the acoustic measurements at the WDR broadcast studios with me at any time of the day. This was rock'n'roll. Another very special thanks to Johannes Arend for intense support with the listening experiments.

What would science be without that terrific scientific-looking hardware? Special thanks for transforming my crazy ideas, professionally sketched on a wrinkled tissue, into solid and functional hardware to the great guys from the mechanic workshop of the Institute of Communications Systems at TH Köln, Aaron Finkenthei, Michael Söntgenrath and Gerd Mies. Your work is fantastic.

Thanks to the large amount of people who participated in the listening experiments. Particularly many thanks to the group of expert listeners who spent a lot of time

and effort. You are all really great. Sorry, I cannot mention all of your names here. Also many thanks to several bachelor and master students who directly or indirectly supported this work.

Thanks to the WDR radio broadcast studios in Cologne and especially Markus Haßler and Benedikt Bitzenhofer for their steady support, and for enabling the acoustic measurements.

Thanks to IOSONO/BARCO Audio Technologies for providing the fantasic binaural renderer that was used for the listening experiments. Very special thanks to Lutz Altmann, who modified the rendering software and remote protocol according to our requirements.

Thanks for many fruitful discussions, scientific exchange, and a lot of joyful evenings at various conferences and meetings to my dear colleagues from Technical University of Berlin and University of Rostock (in alphabetic order) Dr. Jens Ahrens, Fabian Brinkmann, Vera Erbes, Matthias Geier, Michael Horn, Dr. Alexander Lindau, Frank Schultz, and Dr. Hagen Wierstorf.

You cannot write a doctoral thesis without massive support from your family. I deeply thank my parents Monika Bernschütz-Hörnchen and Walter Bernschütz, who both passed away just before I defended and published my work. You did not know it, but this work was dedicated to both of you from the beginning. At the end you could not be here with me to celebrate the graduation. I miss you and thank you so much for everything.

Many special thanks go to my dear relatives in the USA, to Dr. Hans Georg Ritter at Berkeley Lab, for proofreading my work.

Thanks to my sister Dr. Rebecca Schorm-Bernschütz with her husband Dr. Alexander Schorm and my brother Bastian Bernschütz for their support in many different ways. I also thank my second family Ursula and Peter Spang, as well as Dr. Ulrike and Carsten Weinhold for their kind support.

I deeply thank my partner Dr. Astrid Spang for standing by my side during all this time, for her love, patience, and full-time support. Many thanks to our son Armin who enlightened my days during the last few weeks of this work.

Benjamin Bernschütz,
December 2015

Abstract

This thesis discusses a field-related recording technique based on microphone arrays and orthogonal sound field decomposition that delivers a suitable description for dynamic binaural reproduction.

Dynamic binaural reproduction refers to a mostly headphone-based reproduction method that allows for presenting localizable virtual sources and accounts for the head movements of the recipient in order to decouple them from the spatial orientation of the virtual auditory scene. Increased source localization and externalization stability can be regarded as primary advantages compared to classic static binaural reproduction. Spatially stationary or dynamic virtual sound sources can be presented that maintain their spatial positions or move in relation to a fixed external world coordinate system, which is independent from the recipient's head movements.

Dynamic binaural reproduction requires either object-based audio production or specific field-related recording techniques. The focus of this thesis lies on the latter. Using microphone arrays paired with orthogonal sound field decomposition appears to be a particularly promising approach for field-related dynamic binaural recording. It is based on an elegant, closed-form mathematical solution and allows accounting for head-tracking in all rotational degrees-of-freedom during the playback of a recorded auditory scene. Theoretically, even translation of the recipient can be considered. The method inherently comprises individualization capabilities by employing individual head-related transfer functions (HRTFs) and allows for point-to-multipoint distribution. Due to the close mathematical relationship with higher-order Ambisonics (HOA), the respective formats and codecs can be used for storage and distribution of the audio data.

The theoretical mathematical approach under ideal physical conditions is discussed and a closed-form solution is derived. Due to constraints in technical systems, such as discrete spatial sampling or noise in the signal paths for instance, ideal conditions cannot be maintained in practice. The major constraints are pointed out and their specific impact is analyzed and assessed. Various approaches for improving the transmission characteristics of the system are proposed and evaluated. The perceptual properties under dedicated technical constraints and realistic conditions are assessed in listening experiments. Optimal technical parameters for the system are also determined. It turns out that an array-based system for dynamic binaural recording with satisfying perceptual properties can be realized within reasonable technological and economical limits.

Abstract in German

Die vorliegende Dissertationsschrift behandelt ein feldbezogenes räumliches Audioaufnahmeverfahren, das auf Mikrofonarrays und orthogonaler Schallfeldzerlegung beruht und eine geeignete Beschreibung für dynamische binaurale Wiedergabe liefert.

Dynamische binaurale Wiedergabe bezeichnet ein meist kopfhörerbasiertes räumliches Audiowiedergabeverfahren zur Darbietung lokalisierbarer virtueller Schallquellen, das die Kopfbewegung des Rezipienten berücksichtigt, um sie von der räumlichen Orientierung der wiedergegeben virtuellen auditorischen Szene zu entkoppeln. Die wesentlichen Vorteile gegenüber statischer binauraler Wiedergabe bestehen in einer verbesserten Lokalisation und Externalisierung der virtuellen Quellen, sowie der Möglichkeit, statische oder dynamische virtuelle Quellen wiederzugeben, die unabhängig von der Kopfbewegung des Rezipienten ortsfest verbleiben oder sich in Bezug zu einem statischen weltbezogenen Koordinatensystem bewegen.

Dynamische binaurale Wiedergabe setzt entweder objektbasierte Audioproduktion oder spezifische feldbezogene Aufnahmeverfahren voraus. Letztere liegen im Fokus dieser Arbeit. Der Einsatz von Mikrofonarrays in Kombination mit orthogonaler Schallfeldzerlegung stellt hierzu einen vielversprechenden Lösungsansatz dar. Das Verfahren beruht auf einer eleganten mathematisch geschlossenen Lösung. Kopfbewegungen des Rezipienten können in allen rotatorischen Freiheitsgraden berücksichtigt werden. In der Theorie lassen sich auch translatorische Freiheitsgrade einbeziehen. Durch Einsatz individueller kopfbezogener Übertragungsfunktionen (HRTFs) kann die Wiedergabe individualisiert werden. Das Verfahren eignet sich für Punkt-zu-Mehrpunkt Übertragung. Aufgrund enger mathematischer Verwandtschaft zum higher-order Ambisonics (HOA) Verfahren, lassen sich die dort eingesetzten Formate und Codecs zur Speicherung und Übertragung der Audiodaten nutzen.

Unter Annahme idealer physikalischer Bedingungen wird zunächst der theoretische Ansatz diskutiert und eine mathematisch geschlossene Lösung abgeleitet. Aufgrund verschiedener Einschränkungen in technischen Systemen, wie beispielsweise raumdiskrete Abtastung oder Rauschen in den Signalwegen, sind in der Praxis allerdings keine idealen Bedingungen erzielbar. Die wichtigsten Einschränkungen werden aufgezeigt und ihr jeweiliger Einfluss auf das Systemverhalten untersucht. Verschiedene Methoden zur Verbesserung der Übertragungseigenschaften werden diskutiert. In Hörversuchen werden perzeptive Eigenschaften des Systems im Hinblick auf spezifische technische Einschränkungen sowie realistische Bedingungen evaluiert. Ferner werden optimale Systemparameter ermittelt. Es zeigt sich, dass arraybasierte Systeme für die feldbezogene dynamische Binauralaufnahme mit guten perzeptiven Eigenschaften unter vertretbarem technischen und wirtschaftlichen Aufwand realisierbar sind.

Prolog

«Sorry, we're sold out tonight.» This is the final straw. Jon was looking forward to this concert for ages. It's gonna be the very last show of his all-time favorite artist. That's it. Jon is about to turn away. «Hey, you're looking so upset. I'm really sorry. But did you hear they're checking out a crazy new technology, like immense ... or imitative live sound transmission today? They told me it would be incredible – just feels like being here with the fellows at the venue or so. Don't know how it works, but maybe you wanna give it a try.» The doorman passes Jon a flyer. «Okay, thanks, maybe I'll check it out.»

Back at home Jon opens himself a can of beer and remembers the flyer. *Well, might be better than nothing*, he thinks. With expectations to hear one of these radio concert transmissions they do on Friday evenings, he starts to follow the simple instructions on the flyer. Download the application to your mobile phone. Plug in and put on your headphones with tracking sensor. Log in to the stream. Close your eyes. Welcome to the show!

John sits back. After a short while the mobile phone starts grabbing the stream. Jon closes his eyes. A friendly telephone information voice annotates: «Welcome to our new immersive live sound streaming service and welcome to the show tonight. We'll fade you over to the venue in about 5 seconds.»

At the same time Christine made herself comfortable on the plush sofa of her shared apartment in Berlin and follows the stream of the Vienna Philharmonic Orchestra playing tonight. She enjoys this application that achieved bringing her to the most fascinating concert halls all over the globe during the last weeks. Everything started with buying these new pink headphones in June, wearing a special sensor, and the little flyer in the boxing announcing the application. She loves how it sounds here. It seems like being in a particularly good seat tonight. The orchestra spreads wide in front of her and she feels the depth of this wonderful stage, she hears every breath of the pianist behind his piano and the violins are clear like diamonds in front of her. The amazing room acoustics gently envelopes her from all around. What a wonderful evening here in Vienna.

Karl is very pleased about this new teleconferencing system they brought for his company. The last years were exhausting. He didn't even have time to recover from the last jet lag before the next one was around the corner. Finally, he wouldn't have to travel that much anymore. And finally, he could spent more time with his wife and his children. Teleconferencing underwent a severe evolution lately. Thanks to any new kind of audio technology it feels as if he were sitting right at the conference table with his colleagues and business partners in New York this Monday morning. *It's a funny*

piece of technology, he thinks by himself, *they just put something that looks like a soccer ball in my place at the table if I'm not there with them – they told me it was kind of a microphone array or something like that.* Since they brought up the soccer ball and the new headphones with tracking sensor, you can definitely point to the position of every single person sitting at the table around you and the speech is emerging at some distance from you as if you were really there in the conference room. You can even turn your head towards each person who is talking. Karl now is able to take part even in the most complex conversations with many people around the table, where sometimes more than one single track of conversation is going on. All of this was near to impossible with the old system the company had. Complex conversation always was a complete mess. This is the reason why he traveled to any ever so insignificant meeting in any corner of the world.

After 5 seconds suddenly the world around Jon starts to change. He is surrounded by a dense crowd in a huge venue. A girl starts to scream and to whistle right in his back. Completely scared he drops his drink. He opens his eyes and spins around. He continues to hear the whistle and screaming coming from the same place as before, but he does not see anything but his old run-down dresser in the corner of his room. There is nobody. He rips off his phones. Silence. «Okay, what the ... *this* is really kind of scary», he speaks out loud. He becomes aware of the drink leaking to the carpet. He stops the stream and cleans the floor.

Having gotten over the first shock, Jon sits back again and puts his headphones on. He logs in to the stream and closes his eyes. He is back at the venue. Right here. Right now. He hears laughter, screaming, whistles around him. Some guys talking about the afternoon's soccer results right to his left. It feels as if he could touch them standing at short distance, while the tremendous concert hall surrounds him from everywhere. All of a sudden, a thundering noise goes through the venue. The crowd instantly starts to go crazy. The musicians hit the stage. Everybody is screaming and clapping hands with a kick-drum giving the beat. Jon has goosebumps. A shiver runs down his spine. He feels like wanting to jump up from his chair and scream along with his invisible fellows next to him. The show starts. Everybody goes crazy in here. Jon hears the huge PA rigs from above and the backline amps screaming from the stage in front of him. The crowd around him sings along every single line of his favorite song. Jon instantly starts to sing with them. What a night! This is raw. This is live. This is the full packet of emotions that a live concert can deliver. And this is definitely *not* one of these radio concert transmissions they do on Friday evenings.

This thesis is about the technology.

Contents

Abbreviations

ABRIR	array based binaural room impulse response
ADC	analog to digital converter
AN	anchor
ANOVA	analysis of variance
ARIR	array room impulse response
BEM	boundary elements method
BEMA	bandwidth extension for microphone arrays
BRIR	binaural room impulse response
CD	constant directivity
CI	confidence interval
CR7	control room 7 at WDR Cologne
DAC	digital to analog converter
DFT	discrete Fourier transform
DI	directivity index
DOA	direction of arrival
DOF	degree of freedom
DTFT	discrete-time Fourier transform
EOB	effective operational bandwidth
ER	early reflection
FFT	fast Fourier transform
FIR	finite impulse response
GUI	guided user interface
HOA	higher-order Ambisonics
HR	hidden reference
HRIR	head-related impulse response
HRTF	head-related transfer function
IIR	infinite impulse response
ILD	interaural level difference
IR	impulse response
ITD	interaural time difference
LBS	large broadcast studio at WDR Cologne
MIMO	multiple input – multiple output
MTB	motion tracked binaural sound
MUSHRA	multiple stimulus with hidden reference and anchor
PA	public adress
PWD	plane wave decomposition

RBRIR	binaural room impulse response with limited modal resolution
REFC	residual error compensation filter
RHRIR	head-related impulse response with limited modal resolution
RHRTF	head-related transfer function with limited modal resolution
RT60	reverberation time 60
SAQI	spatial audio quality inventory
SBS	small broadcast studio at WDR Cologne
SNR	signal-to-noise ratio
VAE	virtual auditory environment
VAH	virtual artificial head
WDR	Westdeutscher Rundfunk
WNG	white noise gain

1 Introduction

1.1 State of Technology

In order to introduce the topic and to motivate this work, a brief overview of the current state of binaural technology is given in the following. The methods for recording, production and playback are discussed.

1.1.1 Binaural Hearing

Before starting with binaural technology, at least a minimal introduction to binaural hearing is indispensable. Binaural hearing refers to the reception of sound events with two ears and the deduced perception of auditory events or auditory scenes. Sound events are evoked by sound sources that produce a specific physical sound field around the listener, which is determined by deviations of the static air pressure and the particle velocity. The ears evaluate deviations of the air pressure and produce bioelectrical signals that are further evaluated by the central auditory system in the human brain. Since the ears are placed at different spatial locations in the sound field and the head directly interacts with the sound field, i.e. evokes scattering effects, the sound field differs at the two ear canal entrances, except for sound events in certain spatial directions. The actual difference manifests itself as spectral and temporal differences that depend on the direction of sound incidence. These interaural differences are globally referred to as binaural cues and are specifically referred to as interaural level differences (ILDs) and interaural time differences (ITDs). Additionally, reflections from the pinna and torso influence the sound field at the ear canal entrance and evoke a certain filtering of the spectrum that is received by the ears, depending on the direction of sound incidence. The coloration effects due to pinna and torso reflections are referred to as monaural cues. Both, monaural and binaural cues are grouped in the collective term localization cues. The auditory system evaluates these cues for localizing sound sources. The principles and mechanisms of binaural hearing are massively simplified. For a comprehensive introduction and overview of binaural hearing, the reader is referred to (Blauert, 1997), which can be considered the primary reference work on this topic.

1.1.2 Binaural Technology

Binaural technology embraces all technological methods that finally yield the technical reproduction of a sound scene based on two separate ear signals in such a way that the human auditory system can read the necessary information for localizing virtual sound sources in the perceived auditory scene.

Binaural technology is based on the assumption that if a technical system produces the same sound pressures at the ear canal entrances of a recipient as a real source would do, the recipient perceives a virtual source that cannot be distinguished from a real source, refer to (Møller, 1992), for instance. Hence, binaural technology aims at a physically correct reproduction of the sound field at the ear canal entrances.

If we further define (room) reflections as sources, which is feasible from a listener's point of view, it becomes explicit that a recipient can be virtually placed in an arbitrary acoustical environment (e.g. a virtual concert hall) that includes active sound sources and passive room acoustics. At best, the subject feels to be actually present in the reproduced virtual scene.

The term binaural technology mostly refers to reproduction using headphones. Alternative approaches are loudspeaker-based transaural reproduction systems using appropriate cross-talk cancellation filters for generating two separated ear signals close to the ear canal entrances. Transaural reproduction is discussed e.g. by Damaske (1971), Cooper and Bauck (1989), Bauck and Cooper (1992), Bauck and Cooper (1996), Gardner (1997), Menzel et al. (2006), or Lentz (2006).

Binaural reproduction systems always comprise exactly two ear signals fitted for appropriate acoustical reproduction close to the ears, which inherently carry localization cues that can be evaluated by the auditory system for localizing sound sources. This discriminates binaural reproduction systems from other (spatial) audio reproduction systems including the conventional stereophonic headphone-based reproduction.

Note that localization comprises both spatial directions in terms of azimuth and elevation and also in terms of distance. At best, a virtual source that is generated by a binaural system is perceived outside the head (externalized) with dedicated distance to the recipient.

Comprehensive research on binaural technology has been conducted in the past and is still being conducted in the present. There is a vast amount of literature available. Binaural technology, as well as certain inherent aspects of binaural hearing are discussed e.g. in (Møller, 1992), (Gilkey and Anderson, 1997), or (Blauert, 2013). A good overview and a discussion of some recent aspects are provided by Lindau (2014a).

The topics discussed in this thesis build the bridge between field-related descriptions and binaural technology. Field-related description refers to approaches such as wave field synthesis, introduced by Berkhout (1988) (refer to (Spors et al., 2008) for a review), and especially near-field compensated higher-order Ambisonics (Daniel, 2003), which is based on Ambisonics introduced by Gerzon (1973) and others.

1.1.3 Static Binaural Recording

The most simple and straight-forward approach to binaural technology is recording a sound scene using an acoustic dummy head or ear microphones and playing it back through headphones. This method is mostly referred to as binaural recoding. It is well-known in science and also in media production practice for many years. An overview of the history of binaural recording technology is provided in (Paul, 2009).

The advantages of this basic method are manifold. First of all, the realization is very simple and does not require any active signal processing – at least in a most basic setup. Real sound scenes of arbitrary complexity can be recorded, transmitted and reproduced. The method requires only two audio channels from recording to reproduction. Signals can be captured and played back in real-time. The captured signal allows for direct point-to-multipoint distribution, i.e. any amount of recipients can join a binaural transmission. Furthermore, using ear microphones for recording a sound scene, the individual anatomical properties of a single subject can be taken into account.

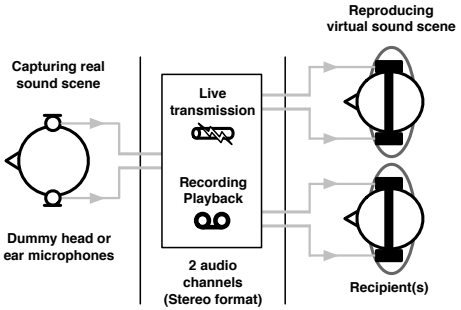


Figure 1.1 Basic system overview of binaural recording.

However, there are some clear restrictions on the other hand. This method basically does not take into account the head movements of the recipient. The auditory scene is always related to the recipient's head and follows his or her head movements during

playback[1]. This differs from the natural experience where the auditory scene is typically independent of the head movements of the recipient and the head can be freely moved in relation to the scene.

Besides of this unnatural behavior of the virtual environment compared to a real environment, the fixed relation between auditory scene and head movement yields a classical problem in binaural recordings. If the head cannot be moved independently from the spatial sound scene, the localization accuracy and stability decrease and front-back confusions for frontal or rearward sources arise. This effect increases if the recordings are non-individual, i.e. do not account for the specific anatomical properties of the recipient. At worst, the entire auditory scene collapses and sources are perceived from straight above or are not even externalized anymore. Several experiments show, approve, and partly explain these phenomena, refer e.g. (Young, 1931), (Wallach, 1940), (Pollack and Rose, 1967), (Perrett and Noble, 1997), (Wightman and Kistler, 1999), (Begault et al., 2001), or (Mackensen, 2004). Similar effects were observed for presenting static binaural recordings and subjects with physically fixed heads exposed to real sound scenes. The common consent is that the auditory system apparently evaluates (small spontaneous) head movements for stabilizing the localization of sources and avoiding front-back confusions. Head rotation in the horizontal plane appears to play a major role herein. The differential information resulting from head rotation is generally referred to as dynamic cues.

From this perspective we understand that static binaural recordings inherently cannot work very well at all, being further prone to massive inter-individual perceptual differences, especially if the recordings are non-individual. Besides the need for wearing headphones for reproduction, this might indeed be one of the major reasons why static binaural recodings did not achieve any true commercial breakthrough in the past.

Even if binaural recording basically allows accounting for the individual anatomical properties of a single recipient (either if the recipient himself wears ear microphones during recording or a specific dummy head with suitable properties is built) and allows point-to-multipoint transmission, it is not possible to combine both features at the same time. There is either individual point-to-point transmission or non-individual point-to-multipoint transmission.

[1] If the scene is recorded by a subject wearing ear microphones, the subject's head movements, which are not identical with the recipient's head movements, are firmly encoded in the recorded signal. This is not discussed here.

1.1.4 Head-Tracking

The key to overcoming many of the restrictions and uncertainties in binaural recordings is tracking the head movements of the recipient and adapting the reproduced signals accordingly, in order to establish dynamic localization cues. Technology for tracking the head orientation or position is state-of-the-art. An overview of different methods is given by Hess (2012).

However, the actual adaptation of the ear signals accounting for the head movements is a technological challenge. The next sections describe some of the currently known approaches to this problem.

1.1.5 Rotating Dummy Head

Listening experiments using a motorized dummy head that is coupled to a head-tracking system were conducted by Mackensen (2004), for example. A motorized dummy-head for free rotation in three degrees of freedom (DOF) is e.g. proposed by Toshima et al. (2003). Similar early approaches using mechanical rotation of a dummy head coupled to the recipient's head were described much earlier by Kock (1950) and others.

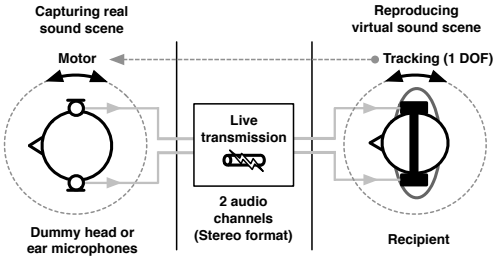


Figure 1.2 Basic system overview with rotating dummy head. The system can be expanded to more degrees of freedom by adding more motors, compare e.g. (Toshima et al., 2003).

This general approach yields clear improvements concerning localization, front-back confusion, or externalization. In practice it has severe limitations. First of all, the method can only be applied for live transmission and does not allow for recording and time-delayed playback. It only allows for point-to-point transmission, since the head can only follow a single recipient. Possible structure-borne noise in mechanical systems or motor reaction times, which might be a question of optimization only, are not even

considered here. All in all, the method might be useful for research purposes, but there hardly appears to be any serious practical application.

1.1.6 Dynamic Binaural Synthesis

The availability of computers and digital signal processing brought up a different approach that is commonly referred to as dynamic binaural synthesis (Wenzel et al., 1990), (Begault, 1994), (Reilly and McGrath, 1995), (Karamustafaoglu et al., 1999).

The influence of pinna, head and torso on the sound transmission from a point source in the free field to the ear canal entrances is determined by a head-related transfer function (HRTF)[2] (Møller, 1992). The latter, in terms of linear system theory, describes a pair of filters with certain specific magnitude and phase responses that vary with the direction of sound incidence. A HRTF describes the pure physical transmission properties from a source to the ear canal entrances without carrying any specific information content. It inherently comprises localization cues that can be evaluated by auditory system for localizing the source position.

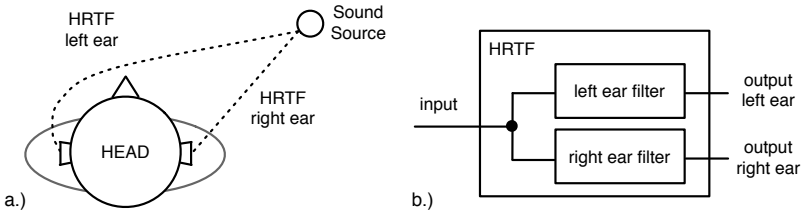


Figure 1.3 a.) Illustration of the head-related transfer function. b.) Simple linear system description of the HRTF with the filter pair for the left and the right ear.

HRTFs can be acquired by acoustical transfer function measurements in an anechoic environment either using a dummy head or ear microphones worn by a subject. Besides acquiring HRTFs by acoustic measurements, their numeric approximation using boundary elements method (BEM) is a promising alternative approach (Katz, 2001a), (Katz, 2001b), (Gumerov et al., 2010).

An extension of the concept of HRIRs are binaural room impulse responses (BRIRs), which can be regarded as a specific case of HRIRs. BRIRs are acquired conducting impulse response or transfer-function measurements in a non-anechoic environment

[2] HRTFs are also referred to as head-related impulse responses (HRIRs), which is an equivalent time-domain representation. Both are related through the time-frequency Fourier transform (Beerends et al., 2003, p 141).

such as a room or concert hall. As an alternative to measuring BRIRs, room simulation tools (Krokstad et al., 1968), (Vorländer, 2008) can be employed for generating artificial BRIRs for instance. While HRIRs only regard the properties of the subject itself, BRIRs integrate the surroundings as well. In addition to the direct sound, BRIRs include early reflections and reverberation. A BRIR describes a fixed constellation of source, room, and receiver.

The actual information content is brought in by convolving the HRTF or BRIR with an anechoic monaural audio signal in a separate step. When playing back the convolution product of HRTF and audio signal through headphones, the recipient perceives an externalized virtual source in free space that reproduces the audio signal. Using a BRIR instead, the recipient perceives an externalized virtual source located in a room and the recipient is enveloped by the room acoustics as if he were present in the room. Hence, the binaural signal is synthesized from a combination of a pure physical description of the transmission path and a dry audio signal.

If both head-tracking, as well as a set of HRTFs or BRIRs for different spatial incidence directions are provided, differently orientated HRTFs or BRIRs can be selected and exchanged dynamically during playback depending on the head movement of the recipient. The process basically operates against the head movement, which maintains the auditory scene fixed in space or, more precisely, unchains it from the head movements of the recipient. From a different perspective, the recipient can move his head independently form the spatial orientation of the sound scene. This is equivalent to reality. As a consequence, the auditory system can access dynamic cues from head movements for stabilizing localization.

Dynamic binaural synthesis is performed using a binaural renderer, which manages the dynamic exchange and convolution of HRTFs or BRIRs with dry source signals. A popular state-of-the-art renderer is provided by Geier et al. (2008) for instance. Current state-of-the-art systems for dynamic binaural synthesis do already achieve highly plausible reproduction of acoustical environments (Lindau, 2014a), partly even touching the edge of authenticity (Brinkmann et al., 2014).

Dynamic binaural synthesis cannot be compared directly to binaural recordings, since it follows a basically different approach. While binaural recordings are based on capturing real existing sound scenes, dynamic binaural synthesis is generally based on composing artificial sound scenes. Hence, no true recordings in a classical sense but only composed productions can be realized. Ideally, the audio signals, e.g. narrators, instruments, sound effects, or background noise, should be recorded in an anechoic environment and can be placed in virtual acoustic environments inside the renderer in a separate

step. This has several different advantages and drawbacks in practice, which are not discussed at this point.

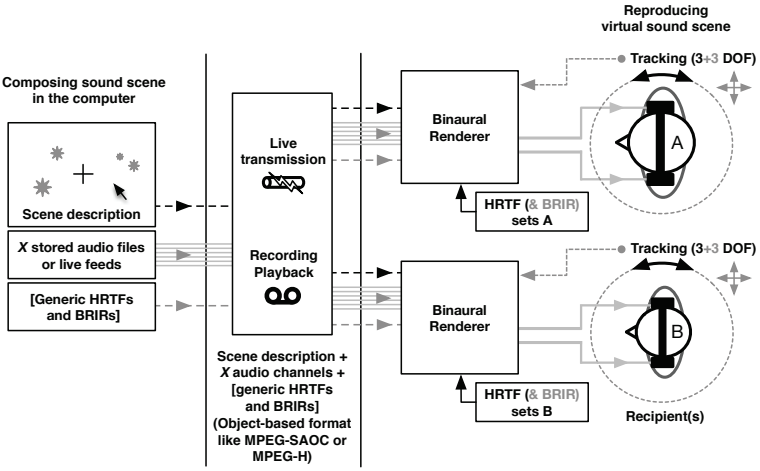


Figure 1.4 Basic system overview for a dynamic binaural synthesis system.

The final product is basically an object-based scene description (Geier et al., 2010) that is rendered in real-time during playback in a binaural rendering instance exclusively assigned to a single recipient. This is indispensable in order to account for the recipient's head rotation and it inherently allows accounting for specific anatomical properties of the listener by using individual HRTFs or BRIRs, refer to Section 3.12.1. Point-to-multipoint transmission is possible. Object-based scene descriptions and the audio source signals can be distributed to multiple recipients. Formats, such as MPEG-H Audio (Herre et al., 2014) or MPEG SAOC (Engdegard et al., 2008), provide first approaches to transmitting object-based audio. An alternative to transmitting object-based audio is performing a static 2-channel downmix that is compatible with the Stereo format, i.e. the signals can be transmitted through stereo infrastructure but must be reproduced using headphones. However, the static downmix yields the identical situation and problems discussed in the scope of static binaural recordings above, since the downmix does not allow for using head-tracking.

All in all, dynamic binaural synthesis has several advantages including head-tracked and individual reproduction, as long as object-based descriptions are used from production to reproduction. Nevertheless, it is a distinct approach and concept that does not

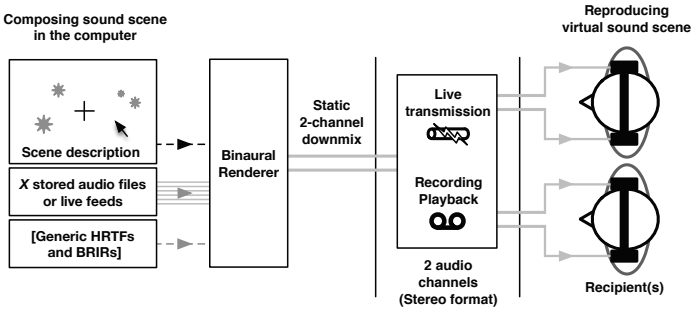


Figure 1.5 Basic system overview for a dynamic binaural synthesis system with static 2-channel downmix for transmission or recording in Stereo format.

replace the classical binaural recording at all, as it does not allow for field-related capturing, transmission, and playback of real existing sound fields.

However, this might be essential in certain situations such as live concert recording or transmission including room-acoustics, artist-room interaction, as well as the indispensable vivid and enveloping ambience noise from the audience. Other examples are teleconferencing, live-transmission of sport events, capturing sound-scapes, or improvisational theater. All of these scenarios actually live from a certain spontaneousness, agility, and rawness in the sound scene, which can hardly be post-produced using a binaural renderer in a satisfying way. As a consequence, a different approach is required, which truly extends the classical binaural recording methods by head-tracking and individualization capabilities.

1.1.7 Motion-tracked Binaural Sound (MTB)

Motion-tracked binaural sound (MTB) was proposed by Algazi et al. (2004). Refinements are described in (Melick et al., 2004), (Algazi et al., 2005), and (Hom et al., 2006). MTB is perceptually evaluated by Lindau and Roos (2010).

MTB is based on the assumption that a rigid sphere of head diameter that is equipped with flush mounted microphones is capable of delivering the most important binaural cues such as ITDs and ILDs. A rigid sphere in the sound field evokes comparable acoustic scattering effects to a human head, especially at lower temporal frequencies, compare Section 4.3.2.2.1. The underlying idea is not basically new and the methods were applied in the scope of static recordings long before the introduction of MTB (Theile, 1986a).

The new aspect of MTB is equipping the rigid sphere with a multitude of microphones along the equator and switching over, fading over, or interpolating between the different signals depending on the head rotation of the recipient. The signals of opposing microphones are directly used for binaural auralization. Hence, MTB is based on a microphone array but does not involve classical array processing such as beamforming or sound field decomposition.

The MTB approach allows for field-related recording or transmission of real existing sound scenes and enables using head-tracking during reproduction. Point-to-multipoint transmission is possible, albeit proprietary data formats and codecs are required, since the format is not compatible with any existing approaches.

There are some basic restrictions. MTB is based on a highly simplified model of the head that e.g. does not provide pinnae. Hence, monaural localization cues are missing. Even if several creative approaches for customization and individualization of MTB signals are proposed by Melick et al. (2004), no true analytic closed-form solution for providing real generic or even individual HRTFs can be achieved.

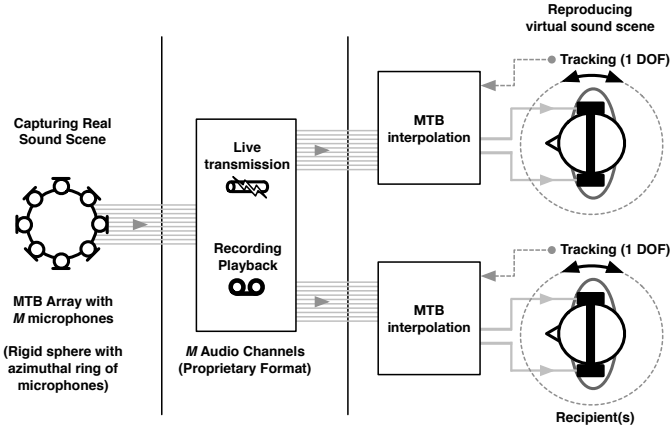


Figure 1.6 Basic system overview for motion-tracked binaural sound (MTB).

Another critical factor is the cross-fading or interpolation between different sensors of the MTB array. Different interpolation methods are proposed in (Algazi et al., 2004), (Melick et al., 2004), and (Hom et al., 2006).

The MTB method, as introduced in the literature, only allows for head-tracking in the horizontal plane, as all microphones are lined up along the equator. The horizontal plane can be considered the most essential degree of freedom. Extending the approach

to other degrees of freedom would be possible if extra microphones at other angles of latitude were added. However, the number of microphones and required transmission channel rises very quickly.

All in all, MTB appears to be a promising approach for practical applications that is comparably easy to realize. On the other hand, it has several restrictions and disadvantages due to the oversimplified head model and does not provide true closed-form solutions for individualization or even using generic HRTFs.

1.1.8 Virtual Artificial Head (VAH)

A different approach for capturing and reproducing field-related descriptions is e.g. proposed by Chen et al. (1992), Mellert and Tohtuyeva (1997), Tohtuyeva and Mellert (1999), Sakamoto et al. (2010), or Rasumow et al. (2011). The approach is entitled virtual artificial head (VAH) by the authors. Refinements, improvements, and further analysis are provided by Rasumow et al. (2013), Rasumow et al. (2014a), and Rasumow et al. (2014b).

The method is based on using microphone arrays and classical delay-and-sum beamforming approaches for approximating the properties of HRTFs. The beam is formed using least square approaches or non-linear cost functions (Rasumow et al., 2013) for achieving HRTF-like characteristics. The limited number of spatially distributed microphones yields a certain deviation or error between the characteristics of a real HRTF and the beamformer.

Even though the array proposed in e.g. Rasumow et al. (2011) is restricted to covering the horizontal plane only, the approach could be extended to resolve arbitrary directions by adding microphones in different height-layers. Restricting to the horizontal plane would not be feasible for capturing real-world sound fields involving room acoustics or sources that are located out of the optimized plane. Sakamoto et al. (2010) follow a similar processing approach but use a spherical rigid sphere microphone array instead, which enables accounting for all spatial directions.

VAH allows for real-time capture, transmission, and binaural reproduction of real existing sound scenes. Point-to-multipoint transmission is possible, albeit proprietary data formats and codecs are required, since the format is not compatible with any existing approaches. Individual HRTFs can be involved for providing highest possible individualization. The method allows full head-tracking capabilities in all rotational degrees, if the microphone array is not restricted to a single plane.

Most of the proposed array designs, delay-and-sum beamforming and least-square approximations, or alternatively proposed methods arise from a purely mathematical or physical background and not from a perceptually motivated perspective or classical

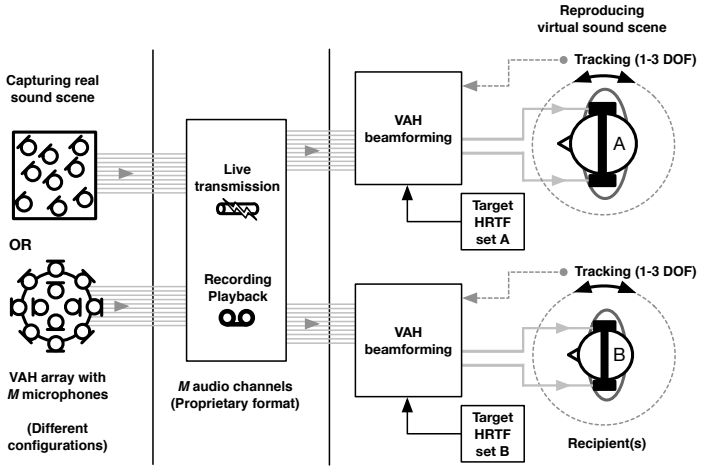


Figure 1.7 Basic system overview for a virtual acoustic dummy-head (VAH).

audio engineering background. Thus, even though in theory, simulations, and first basic localization experiments the approach appears to be very promising, a dedicated evaluation of the perceptual properties has yet to be performed. VAH can be seen as an alternative approach to the methods discussed in this thesis.

1.1.9 Dynamic Binaural Recording (DBR)

At this point we arrive at discussing the approach being subject of the present thesis. For practical reasons, we may simply call it dynamic binaural recording (DBR). The underlying methods are not new and were not invented in this thesis. Before having a closer look at the state of research we start with explaining how it works.

The basic idea consists in capturing the sound field on a spherical surface, decomposing it into orthogonal base functions and performing plane wave decomposition followed by a HRTF-weighted recomposition of the decomposed portions yielding a binaural output signal. Spherical microphone arrays are used for capturing the sound field on the sphere. The method combines graceful closed-form mathematical descriptions with well-proven approaches in audio engineering.

An array-based system for dynamic binaural recording allows for real-time capture, transmission, storage, and binaural reproduction of sound scenes. Point-to-multipoint transmission is possible. Full head-tracking capabilities are provided, not only concerning head rotations in all three degrees-of-freedom but even concerning head translations

in theory. Individual head-related transfer functions (HRTFs) for each single recipient can be involved, yielding the highest possible level of individualization.

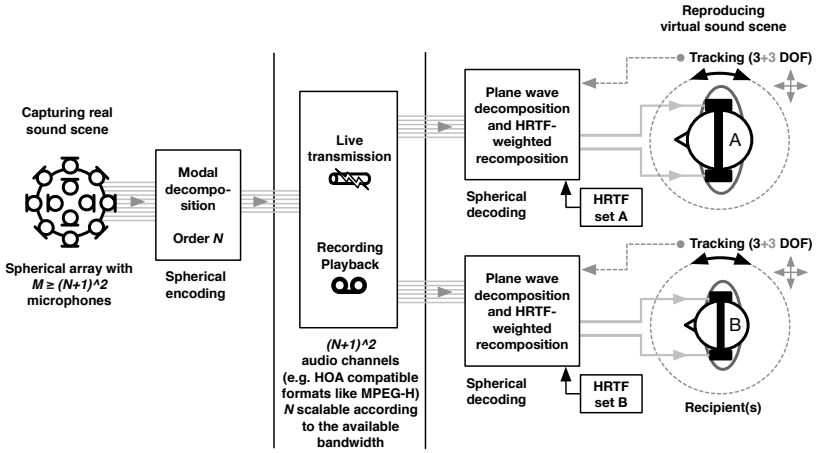


Figure 1.8 Basic system overview for dynamic binaural recording (DBR).

A modal description of the sound field is employed in an intermediate stage between capturing and reproducing the signals, which is based on identical principles and mathematics as the higher-order Ambisonics (HOA) format, refer e.g. to (Gerzon, 1973) or (Daniel et al., 2003). Respective modal sound field descriptions are well-known and their application in audio processing can be considered state-of-the-art. Due to the close relationship to HOA, the signals can be transmitted and stored using available codecs and formats, such as MPEG-H Audio (Herre et al., 2014), that provide HOA support.

In theory, a transparent closed-form solution for binaural auralization can be provided under ideal conditions. However, there are several constraints in technical systems, such as limited modal resolution, spatial aliasing, noise in the transducer paths, as well as several additional error sources yielding non-ideal conditions, that have an impact on the transmission properties. Several of these factors are analyzed and discussed throughout the thesis.

1.2 State of Research and Motivation

The underlying mathematical approach is known as acoustic holography, since it allows for deriving the entire source-free sound field in an interior or exterior volume

from only knowing the sound field on a closed surface. The basic ideas trace back to Berkhout (1988) and Berkhout et al. (1997), who introduced the principles of wave field synthesis and analysis to acoustics and conducted intense research of these topics. They introduced loudspeaker and microphone arrays for reproducing and capturing complex spatial sound fields. Another important milestone was set by Williams (1999), who introduced the principles of Fourier acoustics and introduced what is referred to as spherical acoustics. In the meantime, extensive research was performed by many groups all over the world and uncountable publications emerged.

Amongst others, Meyer and Elko (2002) and Abhayapala and Ward (2002) discuss approaches in spherical phase mode array processing. Similar ideas and circular arrays were introduced before. Rigorous research on spherical microphone arrays and sound field decomposition was performed by Rafaely (2004), who since was involved in about 100 publications on this topic. Much of his work is condensed in (Rafaely, 2015), which is to be released at the time of finishing this thesis. Important contributions, especially in the context of generating binaural signals from array signals, are made by Duraiswami et al. (2005a), who published several works on spherical arrays, sound field decomposition and modal decomposition of HRTFs. Many of the mathematical foundations are described in (Gumerov and Duraiswami, 2004). All in all, spherical microphone arrays and sound field decomposition methods can be considered well-known.

Binaural auralization of modal sound field descriptions captured by spherical microphone arrays is not a new topic either, but the number of contributions on this specific topic is lower. The theoretical approach was first described by Duraiswami et al. (2005b). The authors propose decomposing the sound field using a microphone array and spherical harmonic decomposition techniques for obtaining directional signals and to convolve the latter with with HRTFs for auralization.

Similar approaches are described by O'Donovan et al. (2008), Melchior et al. (2009), Melchior (2011), Shabtai and Rafaely (2012), and Shabtai and Rafaely (2013a), for instance. Spors et al. (2012b) compare modal versus delay-and-sum beamforming approaches for array-based binaural auralization. Spors and Wierstorf (2012) and Salvador Castaneda et al. (2013) analyze different aspects concerning the deviation between HRTFs and the incidence of analytical plane waves to a simulated array-based binaural system. Schultz and Spors (2013), as well as Winter et al. (2014) examine localization properties of an array-based binaural system with special regard to head translation. Shabtai and Rafaely (2013b), as well as Jeffet and Rafaely (2014) describe approaches using array-based binaural auralization for improving speech intelligibility in reverberant spaces. Rettberg and Spors (2013) analyze the impact of noise in array-

based binaural systems. Sheaffer et al. (2014a) propose using the finite difference time domain (FDTD) method for deriving binaural impulse responses from a spherical array response. Sheaffer and Rafaely (2014) and Sheaffer et al. (2014b) propose equalization of binaural room responses derived from spherical array processing in order to minimize the spectral deviation from original binaural room impulse responses.

Basic and partly informal listening experiments with respect to singular aspects are covered in only a few publications. Melchior et al. (2009) and Melchior (2011) perform listening experiments in order to compare different stimuli and array configurations for binaural auralization using a dual-radius spherical microphone array. Jeffet and Rafaely (2014) evaluate the trade-off between binaural reproduction and enhanced spatial selectivity on perception using microphone arrays for binaural auralization in a basic listening experiment. Sheaffer and Rafaely (2014) and Sheaffer et al. (2014b) evaluate equalization strategies in listening experiments.

More profound, widespread and formal perceptual evaluation is described only by two authors. Song et al. (2011) compare a spherical array beamforming approach for binaural auralization versus a dummy head with respect to different aspects like apparent source width, spaciousness, preference, or localization accuracy. Avni et al. (2013) perform listening experiments comparing array based binaural auralization versus original binaural signals with respect to different attributes such as muffled or bright sound, smearing of transients, high frequency artifacts, balance of timbre, accuracy of localization, perceived source distance, and spaciousness. A repertory grid technique is applied to evaluate the results.

Very basic questions concerning the theoretical methods and the impact of dedicated technical constraints on the binaural output signals still remain open from a technical, as well as from a perceptual point of view. The non-trivial and yet not fully clarified theoretical approach, a rather complex and delicate signal processing chain to be implemented, the need for suitable, trustful and consistent measured data-sets, and a specific listening test environment make a valid, comprehensive, and meaningful analysis covering the entire path from theory to evaluating perceptual properties a sophisticated and elaborate challenge. Yet no comprehensive and fully conclusive work is available that embraces anything from theory to perceptual evaluation in methodically consistent steps.

This thesis starts with a theoretical closed-form solution under ideal conditions. The major constraints in technical systems are pointed out, discussed, and analyzed. A comprehensive perceptual evaluation of the system and of dedicated specific aspects concludes the work. Theoretical deduction, as well as the results from simulations and listening experiments yield clear recommendations for the construction of technical

systems in practice. This is the first wide-embracing work that includes formal perceptual evaluation dedicated to microphone arrays and sound field decomposition for dynamic binaural recording. Several new aspects are developed and discussed, such as the modal mismatch of microphone arrays and HRTFs, for instance, which turns out to be one of the major factors of influence concerning array-based binaural recording. A new approach for patching portions of the array response that are disturbed by spatial aliasing is proposed, discussed, and evaluated. The dedicated impact of limiting the modal amplification gain, spatial aliasing or noise in the transducer channels on the binaural output signal is analyzed and evaluated. Some similar analysis was done by other authors in the context of spherical microphone arrays and plane wave decomposition before, but the impact on the binaural output signal using a true closed-form array-based binaural system turns out to be quite different from a single plane wave decomposition output used in other analyses. Those examples show that this thesis introduces several new and relevant aspects to the topic.

1.3 Objectives of the Thesis

The subject of this thesis is spherical microphone arrays and sound field decomposition techniques for dynamic binaural recording. Many of the approaches and considerations are driven rather by a practical or audio engineering point of view rather than by abstract mathematical or physical considerations. Nevertheless, a comprehensive and solid theoretical introduction to the topic is provided in a first step.

The thesis has different major objectives:

- Develop a theoretical closed-from analytic approach under ideal conditions by starting from the principles of wave propagation for deriving a rotatable binaural signal based on the knowledge of the sound field on a spherical surface. The spherical surface is a generalization of spherical microphone arrays and the rotatability of the binaural signal aims at providing head-tracking capabilities during reproduction in practice.
- Point out the major constraints that arise in technical systems and analyze and assess the impact of the main influencing factors.
- Propose and evaluate approaches to optimize the transmission characteristics of the system with respect to dedicated technical constraints.
- Assess the perceptual properties of a technical system for array-based binaural recording under dedicated isolated constraints and under realistic conditions.

- Determine feasible configurations for technical systems in practice, based on the outcomes of simulations and listening experiments.
- Assess whether a technical system with satisfying perceptual properties can be realized within reasonable technological and economical limits.
- Last but not least, an implicit objective is to provide a useful base of theory, software, and data-sets to the scientific community for continuing research.

The listening experiments in this thesis do not aim at assessing the properties of binaural technology by itself or assessing the plausibility or authenticity of virtually reproduced scenes. The latter is covered in different works like (Lindau, 2014a).

The experiments aim at assessing the auditory differences between array-based dynamic binaural recording and dynamic binaural recording based on rotating dummy heads. The latter is assumed as optimum reference, regardless of the differences between the assumed optimum reference and reality. This appears to be the most reasonable approach for isolating the influences and specific perceptual properties of array-based binaural recording.

The entire approach can be described in terms of linear time-invariant system theory, which permits falling back on impulse responses instead of employing specific content-related signals for describing and analyzing the transmission properties. Impulse responses are easier to handle and allow more generalized and exact system analysis than content-related signals. Therefore, processing and analysis are based on impulse responses throughout the entire work. During the listening experiments content-related signals were presented to the participants, which were generated by convolving the final impulse responses with audio content in the last stage.

Even if this work is based on impulse responses, the approaches can be applied to content-related signals as well, with minor exceptions or restrictions. This is of particular importance with regard to systems capable of recording real sound scenes in practice.

1.4 Overview of the Thesis

In the first chapter, a short introduction to binaural technology is given and its current state of technology is outlined. The state of research concerning array-based binaural recording is discussed. The present work is motivated and its dedicated objectives are defined.

In the second chapter, a theoretical approach for dynamic binaural recording based on a sound field description under ideal conditions is discussed. The chapter covers a

detailed description starting from wave propagation in the sound field up to deriving ideal binaural signals based on the knowledge of the sound field properties on an arbitrary sphere around the center of the recipient.

The third chapter discusses constraints of spherical microphone arrays, which impede maintaining ideal properties in practice. Several aspects, such as the binaural reproduction of sound fields with limited modal resolution, radial filter gain limiting, spatial aliasing, uncorrelated noise in the transducer paths, or the surface expansion of the microphone array transducers are discussed and analyzed with regard to array-based binaural recording.

The fourth chapter is dedicated to resources and technology. Since the presented approaches are comparably recent and not yet established in commercial products, there is hardly any appropriate hardware, software or measured data available. Considerable effort was spent in building a respective base of resources and technology from scratch. The chapter e.g. describes the design and construction of a spherical microphone array measurement system, the design, implementation and verification of a sound field analysis toolbox, the acquisition and verification of several data sets, as well as the setup of a suitable environment for performing listening tests.

The fifth chapter describes listening experiments that were performed in order to verify the approach and to assess its perceptual properties. The tests were performed based on both simulated and measured data. Simulated data is suitable for isolating and analyzing the effect of single influencing factors while maintaining ideal conditions apart from the specific factor under test. Measured data, by contrast, is suitable for evaluating the overall performance under realistic conditions. Several factors, such as modal reduction of HRTFs, radial filter amplification limiting, or spatial aliasing are analyzed.

Finally, the approaches, results and conclusion are summarized. The impatient reader may directly refer to this largely self-containing chapter starting on page 240.

«Make things as simple as possible, but not simpler.» A. Einstein

1.5 Nomenclature

In the following, some of the most important conventions, symbols, and notations that are used throughout this work are defined. Additional symbols or deviations from given symbols are explained in the respective context.

Fourier Domains

Be $g(\cdot)$ a function prototype in the time domain and $G(\cdot)$ a function prototype in the frequency domain. Temporal frequencies are generally given as angular frequencies ω. Exceptions are explained in the respective context of occurrence. The six different Fourier domains appearing in this work are denoted as given in the following examples:

$g(t)$	Time domain
$g(\mathbf{x}, t)$	Space-time domain
$\tilde{g}(\mathbf{k}_\mathrm{c}, t)$	Wave spectrum-time domain
$\mathring{g}_{nm}(r_0, t)$	Spherical wave spectrum-time domain
$G(\omega)$	Frequency domain
$G(\mathbf{x}, \omega)$	Space-frequency domain
$\tilde{G}(\mathbf{k}_\mathrm{c}, \omega)$	Wave spectrum-frequency domain
$\mathring{G}_{nm}(r_0, \omega)$	Spherical wave spectrum-frequency domain (Surface expansion coefficients)
$\dot{G}_{nm}(\omega)$	Spherical wave spectrum-frequency domain (Volume expansion coefficients)

Fourier Transforms

$\mathcal{F}_\mathrm{t}\{\cdot\}$	(Forward) time-frequency Fourier transform
$\mathcal{F}_\mathrm{t}^{-1}\{\cdot\}$	Inverse (backward) time-frequency Fourier transform
$\tilde{\mathcal{F}}_\mathbf{x}\{\cdot\}$	(Forward) spatial Fourier transform in Cartesian coordinates
$\tilde{\mathcal{F}}_\mathbf{x}^{-1}\{\cdot\}$	Inverse (backward) spatial Fourier transform in Cartesian coordinates
$\mathring{\mathcal{F}}_{\mathbf{x},nm}\{\cdot\}$	(Forward) spatial Fourier transform in spherical coordinates
$\mathring{\mathcal{F}}_{\mathbf{x},nm^{-1}}\{\cdot\}$	Inverse (backward) spatial Fourier transform in spherical coordinates

Global Proprietary Sub- and Superscripts

$(\cdot)^{\mathrm{A}}$	Coefficients or signals with spatial aliasing artifacts
$(\cdot)_{\mathrm{c,s}}$	Operator, variable or expression specified for Cartesian (c) or spherical coordinates (s)
$(\cdot)_{\mathrm{D,C,Y}}$	Trace of reference: PWD (D), composite (C) or binaural (Y)
$(\cdot)^{\mathrm{l,r}}$	Left ear (l) and right ear (r)
$(\cdot)^{\mathrm{L}}$	Coefficients or signals with limited radial filter amplification
$(\cdot)^{\mathrm{N}}$	Coefficients or signals with noise only at the array inputs
$(\cdot)^{\mathrm{N'}}$	Coefficients or signals with HRTFs replaced by noise of equivalent power spectral density
$(\cdot)^{\mathrm{OS,OSC,RS}}$	Expression specified for open sphere arrays with pressure transducers (OS), open sphere arrays with cardioid transducers (OSC), or rigid sphere arrays with pressure transducers (RS)
$(\cdot)^{\mathcal{R}}$	Rotated function or coefficient set

Angles

In formulas and written text, angles are given in radians. Angles in figures are given in degrees for more comprehensive illustration.

α, β, γ	Rotation angles about coordinate axis
γ_{t}	Aperture angle of the spherical cap or expanded transducer
θ, ϕ	Elevation and azimuth angle (spherical coordinates)
$\theta_{\mathrm{d}}, \phi_{\mathrm{d}}$	PWD steering direction (elevation and azimuth)
$\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}$	Angle of the composite grid node g_{cg}
$\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}}$	Angle of the sampling grid node g_{sg}
θ_j, ϕ_j	Error evaluation angles, elevation and azimuth
θ_k, ϕ_k	Incidence direction of the reflection/wave with index k (elevation and azimuth)
$\theta_{\mathrm{w}}, \phi_{\mathrm{w}}$	Wave incidence direction (elevation and azimuth)
Ω	Solid angle
Ω_{h}	Subject's head rotation angle
Ω_k	Incidence direction of the reflection/wave with index k

Common Mathematical Functions, Operators and Constants

$\lvert\cdot\rvert$	Absolute value
$(\cdot)^*$	Complex conjugate
$(\cdot)!$	Factorial
$\cos(\cdot)$	Cosine
d	Derivative
$d^n_{mm'}(\beta)$	Wigner-d function
$D^n_{mm'}(\alpha,\beta,\gamma)$	Wigner-D function
$D_m(\alpha)$	Reduced Wigner-D function (Euler rotation)
∂	Partial derivative
$\delta(\cdot)$	Dirac delta
$\delta_{ll'}$	Kronecker delta
$\mathrm{e}^{(\cdot)}$	Exponential function
$h^{(1)}_n(\cdot)$	Spherical Hankel function of the first kind
$h^{(2)}_n(\cdot)$	Spherical Hankel function of the second kind
i	Imaginary unit
$\Im\{\cdot\}$	Imaginary part
$j_n(\cdot)$	Spherical Bessel function of the first kind
∇	Del operator (Nabla)
∇^2	Laplacian
$P_n(\cdot)$	Legendre polynomials
$P^m_n(\cdot)$	Associated Legendre functions
π	Mathematical constant for the circle, $\pi \approx 3.14159$
$Q^m_n(\cdot)$	Associated Legendre functions of the second kind
$\Re\{\cdot\}$	Real part
$\sin(\cdot)$	Sine
$\tan(\cdot)$	Tangent
$(\cdot)^{\mathrm{T}}$	Transposition
$y_n(\cdot)$	Spherical Bessel function of the second kind
$Y^m_n(\theta,\phi)$	Surface spherical harmonics

Vectors

$\vec{\mathbf{e}}_{\mathrm{r},\theta,\phi}$	Unit vectors in Spherical coordinates
$\vec{\mathbf{e}}_{\mathrm{x,y,z}}$	Unit vectors in Cartesian coordinates
$\vec{\mathbf{f}}$	Vector field
$\mathbf{k}$	Wave vector
$\vec{\mathbf{n}}$	Normal vector
$\mathbf{x}$	Position vector

Variables

$\hat{a}$	Radial filter amplification limit (linear)
$\hat{a}_{\mathrm{dB}}$	Radial filter amplification limit (dB)
c	Speed of sound in m/s
d_0	Diameter of the surface S_0 (array measurement diameter)
d_{t}	Transducer diameter in m
DI	Directivity index
η_{g}	Sampling efficiency of the grid
f	Frequency in Hz
f_{A}	Spatial aliasing frequency in Hz
f_{s}	Temporal sampling frequency in s^{-1}
g_{cg}	Index of the composite grid node
g_{sg}	Index of the sampling grid node
J	Number of error evaluation angles
k	Wave number
k	Index of the early reflection
$k_{\mathrm{x,y,z}}$	Wave numbers along the x, y and z axis of a Cartesian coordinate system
K	Total number of early reflections
m	Mode
M_{sg}	Number of nodes (sensors/microphones) in the sampling grid
M_{cg}	Number of nodes in the composite grid
n	Order
N	(Highest) order of the system or PWD
N_{sg}	Order of the sampling grid

N_{cg}	Order of the composite grid
ω	Angular frequency in s^{-1}
ω_{i}	BEMA spatial image extraction frequency
$\widehat{p}, \widehat{P}$	Sound pressure amplitude ($\widehat{p} = \widehat{P}$)
Ψ, χ, ζ	Arbitrary constants
r	Radius
r_0	Radius of the surface S_0 (array measurement radius)
r_{RS}	Radius of the rigid sphere
t	Time in s
$\overline{w}_{\mathrm{cg}}$	Mean composite grid weight
x, y, z	Cartesian coordinates

Surface and Volume Descriptors

S	Spherical surface
S_0	Spherical surface with radius r_0
V_{e}	Exterior volume
V_{i}	Interior volume

Specific Proprietary Functions, Operators, Weights, Signals and Coefficients

$A(\omega)$	Isolated spatial aliasing signal from a plane wave impact
$b^{\mathrm{l,r}}(t)$	BRIR
$\mathring{B}_{nm}(\omega)$	Modal beamforming coefficients
$C(\omega)$	Composite signal using discrete sampling nodes in the space-frequency domain
$C'(\omega)$	Ideal composite signal in the space-frequency domain
$C_0(\omega)$	Omni-directional signal at the center of S_0 (center of the array) in the space-frequency domain
$d_n\left(\frac{\omega}{c} r_0\right)$	Radial filters
$\bar{d}_n\left(\frac{\omega}{c} r_0\right)$	Radial filters with soft-knee amplitude limiting
$\tilde{d}_n\left(\frac{\omega}{c} r_0\right)$	Non-critical radial filters
$D(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)$	Output signal of the PWD
$D^{\mathrm{B}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)$	Output signal of a modal beamformer
$\check{D}^{\mathrm{N}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)$	$\overline{w}_{\mathrm{cg}}$-weighted PWD output

$\Delta_{\mathrm{E}}(\theta_j, \phi_j, N, N_{\mathrm{cg}})$	Mean spectral deviation between RHRTF and its related HRTF
$\widehat{\Delta}_{\mathrm{E}}(N, N_{\mathrm{cg}})$	Mean spectral deviation between RHRTFs and its related HRTFs averaged over the entire surface S
$\gamma(t), \Gamma(\omega)$	Gaussian white noise ($\sigma^2 = 1$)
$\bar{\gamma}(t), \bar{\Gamma}(\omega)$	Averaged reference noise
$\gamma_g(t), \Gamma_g(\omega)$	Gaussian white noise realization at grid node g
$\Gamma^{\mathrm{H}}_{g\mathrm{c}}(\omega)$	Gaussian white noise at the composite grid nodes with equivalent power spectral density to HRTFs
$\mathring{\Gamma}_{nm}(\omega)$	Gaussian white noise expansion coefficients
$h^{\mathrm{l,r}}(\theta, \phi, t)$	Head-related impulse response (HRIR)
$H^{\mathrm{l,r}}(\theta, \phi, \omega)$	Head-related transfer function (HRTF)
$\tilde{H}^{\mathrm{l,r}}(\theta, \phi, \omega)$	Interpolated HRTF
$H^{\mathrm{l,r}}_N(\theta, \phi, \omega)$	HRTF with reduced modal order N (RHRTF)
$H^{\mathrm{l,r\ DF}}(\omega)$	Diffuse field HRTF response
$\mathring{H}^{\mathrm{l,r}}_{nm}(\omega)$	HRTF expansion coefficients
$\dot{I}_{nm}$	BEMA spatial image coefficients weighted with $C_0(\omega_{\mathrm{i}})$
$\dot{I}'_{nm}$	BEMA spatial image coefficients
$p(\mathbf{x}, t), P(\mathbf{x}, \omega)$	Sound pressure
p_n	Signal power at order n
$\hat{p}_k$	Amplitude of the reflection/wave with index k
$P_{\mathrm{pw}(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})}(\mathbf{x}, \omega)$	Pressure spectrum of a plane wave arriving from $(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})$
$P_{\mathrm{sw}(\mathbf{x}')}(\mathbf{x}, \omega)$	Pressure spectrum of a spherical wave excited by a monopole source located at $\mathbf{x}'$
$\mathring{P}_{nm}(r_0, \omega)$	Surface expansion coefficients of the sound pressure
$\dot{P}_{nm}(\omega)$	Volume expansion coefficients of the sound pressure
$\dot{P}^{\mathrm{i}}_{nm}(\omega)$	Interior expansion coefficients of the sound pressure
$\dot{P}^{\mathrm{e}}_{nm}(\omega)$	Exterior expansion coefficients of the sound pressure
$\mathring{P}_{nm\ \mathrm{pw}(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})}(r_0, \omega)$	Pressure spectrum surface expansion coefficients of a plane wave arriving from $(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})$
$R(r)$	Radial Solution to the separated Helmholtz equation
$\mathcal{R}_{zyz}(\alpha, \beta, \gamma)$	Rotation operator (Index: Axis)
$\mathring{\mathcal{R}}_{zyz}(\alpha, \beta, \gamma)$	Rotation operator in the spherical wave spectrum domain (Index: Axis)
$T(\gamma_t)$	Spherical cap

$\mathring{T}_{nm}(\gamma_t)$	Expansion coefficients for a spherical cap
$\mathring{T}'_{nm}$	Expansion coefficients for a spherical cap with $\gamma_t \to 0$
τ_k	Time-shift of the reflection/wave with index k
$\Theta(\theta), \Phi(\phi)$	Angular Solutions to the separated Helmholtz equation
$w_{g_{\mathrm{sg}}}$	Quadrature weight of the sampling grid node g_{sg}
$w_{g_{\mathrm{cg}}}$	Quadrature weight of the composite grid node g_{cg}
$\mathrm{WNG}(\omega)$	White noise gain in dB
$WNG^{-1}(\omega)$	Noise amplification in dB
$Y^{\mathrm{l,r}}(\omega)$	Binaural output signal

HRTF and BRIR Angle Conventions

The angles that are denoted in binaural room impulse responses (BRIRs) describe the orientation of the head in relation to a fix external world coordinate system. The angles that are denoted in head-related transfer functions (HRTFs), by contrast, describe the position of the virtual source or the direction of sound incidence. The reference coordinate system is related to the head in the latter case. All HRTFs refer to the far-field throughout this work; no explicit radius is denoted and the given angles describe the virtual source direction or the direction of sound incidence only, instead of a closer determined position of the source in space.

2 Theory of Sound Field Decomposition

2.1 Homogeneous Acoustic Wave Equation

The propagation of a sound wave in a source-free fluid medium is described by the homogeneous acoustic wave equation (Feynman et al., 2011), (Blackstock, 2000), (Möser, 2007)

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = 0, \tag{2.1}$$

which is a classic second-order linear partial differential equation that in the given denotation refers to the variation of the sound pressure in dependence of time and space. The variation of sound pressure over the static air pressure is denoted as p, the time as t, the position vector in a three-dimensional space as $\mathbf{x}$, and c stands for the speed of sound. The symbol ∇^2 represents the Laplacian and ∂ describes the partial derivative. An analogous equation describes the velocity. Dedicated derivations of the wave equation are given e.g. by Feynman et al. (2011), Blackstock (2000), or Möser (2007). The wave equation for sound propagation is derived from the conservation equations for fluid media, i.e. the conservation of mass and the momentum equation (Blackstock, 2000, pp 27–35). The wave equation in general traces back to the 18th century and is based on fundamental work of Isaac Newton, Leonhard Euler, Daniel Bernoulli, Joseph-Louis Lagrange, and others. Every sound wave satisfies the wave equation as long as certain conditions are maintained. In linear acoustics, a complex sound field can be described by the superposition of single sound waves. As a consequence, arbitrarily complex sound fields follow the wave equation likewise. This property for linear systems and equations is referred to as the principle of superposition (Feynman et al., 2011). The linear and lossless acoustic wave equation as denoted above in Eq. (2.1) demands that specific conditions be fulfilled (Möser, 2007), (Spors, 2006). The pressure and density perturbations caused by the wave must be small compared to the static pressure and density for a valid bias-point linearization. Otherwise, methods from non-linear acoustics need to be considered instead. There are several conditions imposed on the propagation medium. The medium must be characterizable as ideal gas with adiabatic state changes (Feynman et al., 2011), i.e. without considerable heat conduction and resultant propagation loss. Furthermore, the medium must be homogeneous and quiescent in order to assure that the relevant parameters are time-

and space-invariant (Spors, 2006). For the propagation of acoustic waves in air with moderate sound pressure levels the conditions are considered to be adequately fulfilled.

2.2 Wave Equation in Cartesian Coordinates

The wave equation, Eq. 2.1, and its related problems can be formulated for different coordinate systems. The choice of the appropriate coordinate system depends on the respective problem set to be solved. In the following, the basic theory is initially discussed for the Cartesian coordinate system, see Figure 2.1, as this is most comprehensible. Later, the theory is transferred to spherical coordinate systems, as this thesis deals with spherical apertures and geometries.

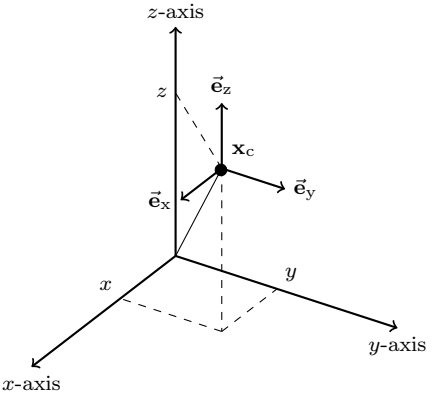


Figure 2.1 Illustration of the $\mathbf{x}_c$ vector in a cartesian coordinate system used in this thesis; $\vec{\mathbf{e}}_\mathrm{x}$, $\vec{\mathbf{e}}_\mathrm{y}$ and $\vec{\mathbf{e}}_\mathrm{z}$ indicate the corresponding unit vectors.

In order to adapt the wave equation Eq. (2.1) to a specific coordinate system, the position vector $\mathbf{x}$ and the Laplacian have to be customized. The position vector in Cartesian coordinates is denoted by

$$\mathbf{x}_c = x\,\vec{\mathbf{e}}_\mathrm{x} + y\,\vec{\mathbf{e}}_\mathrm{y} + z\,\vec{\mathbf{e}}_\mathrm{z}, \tag{2.2}$$

where $\vec{\mathbf{e}}_\mathrm{x}$, $\vec{\mathbf{e}}_\mathrm{y}$ and $\vec{\mathbf{e}}_\mathrm{z}$ describe the unit vectors in the Cartesian coordinate system, see Figure 2.1. The gradient of a function $f(x, y, z)$ in Cartesian coordinates reads

$$\nabla_\mathrm{c} f = \frac{\partial f}{\partial x}\,\vec{\mathbf{e}}_\mathrm{x} + \frac{\partial f}{\partial y}\,\vec{\mathbf{e}}_\mathrm{y} + \frac{\partial f}{\partial z}\,\vec{\mathbf{e}}_\mathrm{z}, \tag{2.3}$$

and the divergence of a vector field $\vec{\mathbf{f}}$ is given by

$$\nabla_{\mathrm{c}} \cdot \vec{\mathbf{f}} = \frac{\partial f_{\mathrm{x}}}{\partial x} + \frac{\partial f_{\mathrm{y}}}{\partial y} + \frac{\partial f_{\mathrm{z}}}{\partial z}, \tag{2.4}$$

whereas

$$\vec{\mathbf{f}} = \begin{bmatrix} f_{\mathrm{x}}(x,y,z) \\ f_{\mathrm{y}}(x,y,z) \\ f_{\mathrm{z}}(x,y,z) \end{bmatrix}. \tag{2.5}$$

The Laplacian is defined as divergence of the gradient of a function $f(x,y,z)$ in Cartesian coordinates written as

$$\nabla_{\mathrm{c}}^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}. \tag{2.6}$$

Hence, the homogeneous linear wave equation Eq. (2.1) in Cartesian coordinates explicitly yields

$$\frac{\partial^2}{\partial x^2}\, p(\mathbf{x}_{\mathrm{c}}, t) + \frac{\partial^2}{\partial y^2}\, p(\mathbf{x}_{\mathrm{c}}, t) + \frac{\partial^2}{\partial z^2}\, p(\mathbf{x}_{\mathrm{c}}, t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\, p(\mathbf{x}_{\mathrm{c}}, t) = 0. \tag{2.7}$$

2.3 Solutions to the Wave Equation in Cartesian Coordinates

The wave equation Eq. (2.1) is solved by an arbitrary function that is twice differentiable (Kuttruff, 2004, p 47). The most general solution for an l-th order linear partial differential equation contains l arbitrary functions (Blackstock, 2000, p 8). As the wave equation is a second order linear partial differential equation, the most general solution $p(\mathbf{x}, t)$ is expected to incorporate two functions f_1 and f_2 that are weighted by two constants $\widehat{p}_1$ and $\widehat{p}_2$ yielding a general expression describing plane waves in a three-dimensional space like

$$p(\mathbf{x}_{\mathrm{c}}, t) = \widehat{p}_1\, f_1(-\vec{\mathbf{n}}_{\mathrm{c}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c}} \pm ct) + \widehat{p}_2\, f_2(-\vec{\mathbf{n}}_{\mathrm{c}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c}} \pm ct). \tag{2.8}$$

Depending on the signs of the last term in brackets this expression describes two waves traveling in the same direction or in opposite directions. The latter expression combined with suitable boundary conditions also includes the description of a standing wave. The negative sign of the first term is chosen by convention. The normal vector

$\vec{\mathbf{n}} = [n_\mathrm{x}\, n_\mathrm{y}\, n_\mathrm{z}]^\mathrm{T}$ with $|\vec{\mathbf{n}}| = 1$ points in the direction of propagation and is orthonormal to the isophasic planes. The shape of the waves is determined by the functions f_1 and f_2. Assuming that the wave involves harmonic oscillations only, it is feasible to replace the general functions f_1 and f_2 by sinoidal functions such as sine, cosine or complex exponential functions. Furthermore, for the description of certain specific problems, i.e. a single plane wave propagating in free space, it is appropriate to reduce the general expression that embraces two functions to a single function only. Hence, choosing the complex exponential function to describe a harmonic plane wave in free field conditions, a potential solution with freely chosen signs and an amplitude factor $\widehat{p}$ is given by

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\, \mathrm{e}^{-\mathrm{i}\frac{2\pi}{\lambda}(\vec{\mathbf{n}}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - ct)}, \tag{2.9}$$

where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit. By introduction of a factor $1/\lambda$ the inner argument is being referred to as wavelength λ. The factor 2π is introduced to stretch one wavelength unit to a full circle run and represents one full period of a sinoidal function. The introduced normalization factor accounts for the periodicity properties of harmonic waves. The normalized argument subsequently indicates the number of wave oscillations per full circle run. The normalization inherently introduces the temporal frequency to the argument, which is one of the classical parameters of harmonic oscillations. The wavelength λ and the frequency f are related by $\lambda = c/f$. By taking the latter into account, as well as the relation $\omega = 2\,\pi f$, the factor $2\pi/\lambda$ can be rewritten as ω/c. This quotient is often replaced by the wavenumber k. The dispersion relation associates the temporal frequency and the wave number, whereas sound propagation in air as an approximately ideal gas is assumed to be non-dispersive, i.e. the speed of sound does not considerably depend on the frequency. The dispersion relation reads

$$k = \frac{\omega}{c}. \tag{2.10}$$

Taking the last considerations into account, Eq. (2.9) is now rewritten as

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\, \mathrm{e}^{-\mathrm{i}(k\,\vec{\mathbf{n}}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)}. \tag{2.11}$$

The scalar wave number k appearing in the equation is firmly associated with the direction of wave propagation and is only valid for this particular perspective. For waves or wave fields propagating in ν-dimensional spaces, $\nu \in \{2, 3, ...\}$, it is rather

convenient to pass over to a wave vector that embraces single components according to the base vectors of the respective coordinate system. The transition to a wave vector can be accomplished by melting the scalar wave number k and the normal vector of propagation $\vec{\mathbf{n}}$ to the wave vector $\mathbf{k} = k\,\vec{\mathbf{n}}$. An explicit description for the wave vector in Cartesian coordinates reads $\mathbf{k}_\mathrm{c} = k\,\vec{\mathbf{n}}_\mathrm{c}$. The components of this particular wave vector are given by $\mathbf{k}_\mathrm{c} = [k_\mathrm{x}\, k_\mathrm{y}\, k_\mathrm{z}]^\mathrm{T}$, where k_x, k_y, and k_z describe the trace wave numbers along the axes of the Cartesian coordinate system. Eq. (2.11) is rewritten as

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)}. \tag{2.12}$$

When inserting Eq. (2.12) into the wave equation Eq. (2.1) it becomes apparent that the components of the wave vector are not independent. Applying the Laplacian to Eq. (2.12) in Cartesian coordinates explicitly yields

$$\begin{aligned}\nabla_\mathrm{c}^2\, p(\mathbf{x}_\mathrm{c}, t) &= \widehat{p}\,(-k_\mathrm{x}^2\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)} - k_\mathrm{y}^2\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)} - k_\mathrm{z}^2\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)}) \\ &\Leftrightarrow (-k_\mathrm{x}^2 - k_\mathrm{y}^2 - k_\mathrm{z}^2)\,\widehat{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)}\,.\end{aligned} \tag{2.13}$$

The second derivative of Eq. (2.12) with respect to the time is given by

$$\frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) = -\omega^2\,\widehat{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)}. \tag{2.14}$$

Both of these terms inserted into Eq. (2.1) reads

$$(-k_\mathrm{x}^2 - k_\mathrm{y}^2 - k_\mathrm{z}^2)\,\widehat{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)} - \frac{1}{c^2}(-\omega^2)\,\widehat{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega t)} = 0\,. \tag{2.15}$$

The essential outcome of this procedure is

$$\left(\frac{\omega}{c}\right)^2 = k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2\,, \tag{2.16}$$

which, using the dispersion relation given in Eq. (2.10), finally yields

$$\begin{aligned} k^2 &= k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2 \\ \Leftrightarrow \quad k &= \pm\sqrt{k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2}\,. \end{aligned} \tag{2.17}$$

This equation points out two important relations. First, the scalar wavenumber k corresponds to the length of the wave vector according to

$$k = |\mathbf{k}| = \sqrt{k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2}, \tag{2.18}$$

where the negative branch of the square root is intentionally discarded as the wave number, as well as any vector lengths, are always positive numbers. Second, Eq. (2.17) demonstrates the dependency of the wave vector components when rearranging the equation according to the single vector components yielding

$$k_\mathrm{x} = \pm\sqrt{k^2 - k_\mathrm{y}^2 - k_\mathrm{z}^2}, \text{ for } (k_\mathrm{y}^2 + k_\mathrm{z}^2) \leq k, \tag{2.19}$$

$$k_\mathrm{y} = \pm\sqrt{k^2 - k_\mathrm{x}^2 - k_\mathrm{z}^2}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{z}^2) \leq k, \text{ and} \tag{2.20}$$

$$k_\mathrm{z} = \pm\sqrt{k^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{y}^2) \leq k. \tag{2.21}$$

Two of the three components can be freely chosen. The third one is dependent, since the wave number is constant. When not restricting the value range of the wave vector components, the argument of the square roots in Eq. (2.19) to Eq. (2.21) can become negative entailing imaginary wave vector components like

$$k_\mathrm{x} = \pm\,\mathrm{i}\sqrt{k_\mathrm{y}^2 + k_\mathrm{z}^2 - k^2}, \text{ for } (k_\mathrm{y}^2 + k_\mathrm{z}^2) > k, \tag{2.22}$$

$$k_\mathrm{y} = \pm\,\mathrm{i}\sqrt{k_\mathrm{x}^2 + k_\mathrm{z}^2 - k^2}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{z}^2) > k, \text{ and} \tag{2.23}$$

$$k_\mathrm{z} = \pm\,\mathrm{i}\sqrt{k_\mathrm{x}^2 + k_\mathrm{y}^2 - k^2}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{y}^2) > k. \tag{2.24}$$

Even when employing imaginary wave vector components, Eq. (2.12) formally satisfies the wave equation. The equation in this case describes evanescent waves (Williams, 1999, pp 24–26) that are characterized by an exponential decay of the amplitude in the direction of the imaginary wave vector component. This can be verified by successively inserting Eq. (2.22) to (2.24) into Eq. (2.12) resulting in

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\,\mathrm{e}^{-|k_\mathrm{x}\,x|}\,\mathrm{e}^{-\mathrm{i}(k_\mathrm{y}\,y + k_\mathrm{z}\,z - \omega t)}, \text{ for } (k_\mathrm{y}^2 + k_\mathrm{z}^2) > k, \tag{2.25}$$

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\,\mathrm{e}^{-|k_\mathrm{y}\,y|}\,\mathrm{e}^{-\mathrm{i}(k_\mathrm{x}\,x + k_\mathrm{z}\,z - \omega t)}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{z}^2) > k, \text{ and} \tag{2.26}$$

$$p(\mathbf{x}_\mathrm{c}, t) = \widehat{p}\,\mathrm{e}^{-|k_\mathrm{z}\,z|}\,\mathrm{e}^{-\mathrm{i}(k_\mathrm{x}\,x + k_\mathrm{y}\,y - \omega t)}, \text{ for } (k_\mathrm{x}^2 + k_\mathrm{y}^2) > k. \tag{2.27}$$

In order to obtain a decaying character, the argument of the real exponential must be chosen to have an overall negative sign. A positive sign in the exponential would describe waves with an exponentially increasing amplitude over the distance, which has no physical sense. Evanescent waves occur e.g. in the context of wave transition between different media, when the incident angle exceeds the critical angle yielding total internal reflection of a plane wave (de Fornel, 2001). Evanescent waves also occur in the context of sound radiation from plates (Williams, 1999). Evanescent waves play an important role in the near field and as their amplitude decays exponentially they can be neglected in the far field. All considerations and experiments presented in this thesis refer to the far field. Thus, evanescent waves are not discussed further.

2.4 Fourier Transforms

The Fourier transform is a linear transform that is used to express a function or signal in terms of a sum of different (simpler) functions or signals. The latter are called base functions of the Fourier transform and need to fulfill the orthogonality criterion. The Fourier transform traces back to Jean Baptiste Joseph Fourier (Fourier, 1822) and is of utmost importance for all fields of physics and engineering. The Fourier transform is often used for transforming signals from time domain into the frequency domain representation and vice versa. This for instance enables spectral analysis of time signals. The general concept of the Fourier transform is by far more versatile. For this thesis three types of the Fourier transform are of great relevance. The first one is the classical time-frequency Fourier transform $\mathcal{F}_\mathrm{t}$ using complex exponential base functions. It transforms a function or signal from the time domain to the frequency domain. The two other transforms are spatial Fourier transforms and their use depends on the respective coordinate system. The first spatial transform $\tilde{\mathcal{F}}_\mathrm{s}$ is specified in Cartesian coordinates and transforms a function or signal from the space domain to the wave spectrum domain by employing complex exponentials as base functions once more. The second spatial transform $\mathring{\mathcal{F}}_\mathrm{s}$ is specified for spherical coordinate systems and transforms a function or signal on a spherical surface to the spherical wave spectrum domain. It uses spherical harmonics as base functions. The latter transform will be discussed later, after introducing the spherical coordinate system properties and base functions. The time-frequency Fourier transform $\mathcal{F}_\mathrm{t}$ is defined as (Beerends et al., 2003, p 141)

$$G(\omega) = \mathcal{F}_\mathrm{t}\{g(t)\} = \int_{-\infty}^{\infty} g(t)\, \mathrm{e}^{-\mathrm{i}\,\omega\, t}\, \mathrm{d}t, \tag{2.28}$$

and the inverse Fourier transform from frequency to time domain as (Beerends et al., 2003, p 164)

$$g(t) = \mathcal{F}_{\mathrm{t}}^{-1}\{G(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)\, \mathrm{e}^{\mathrm{i}\,\omega\, t}\, \mathrm{d}\omega, \tag{2.29}$$

where $g(t)$ is a function depending on the time t and $G(\omega)$ a function depending on the angular frequency ω; the latter is associated with the frequency f by the expression $\omega = 2\pi f$. The one-dimensional spatial Fourier transform in Cartesian coordinates (Williams, 1999, pp 1–2) reads

$$\tilde{G}(k_{\mathrm{x}}) = \tilde{\mathcal{F}}_{\mathrm{x}}\{G(x)\} = \int_{-\infty}^{\infty} G(x)\, \mathrm{e}^{\mathrm{i}k_{\mathrm{x}}x}\mathrm{d}x, \tag{2.30}$$

where G(x) denotes a function in the space domain and $\tilde{G}(k_{\mathrm{x}})$ a function in the wave spectrum domain. This transform is very similar to the time-frequency transform with the exception of the sign of the exponent and the variables referring to the respective source and target domains. The corresponding inverse transform is defined by

$$G(x) = \tilde{\mathcal{F}}_{\mathrm{x}}^{-1}\{G(k_{\mathrm{x}})\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{G}(k_{\mathrm{x}})\, \mathrm{e}^{-\mathrm{i}k_{\mathrm{x}}x}\mathrm{d}k_{\mathrm{x}}, \tag{2.31}$$

The variables x and k_{x} are still congruent to the previously given definition, where x denotes the x-component of the position vector and k_{x} the k_{x}-component of the wave vector. Both could alternatively be understood as new independent variables. The first interpretation directly entails the need for a more general expression that involves all vector components, not just a singular one. This expression is obtained by superposing the transforms for the single components yielding

$$\tilde{G}(\mathbf{k}_{\mathrm{c}}) = \tilde{\mathcal{F}}_{\mathbf{x}}\{G(\mathbf{x}_{\mathrm{c}})\} = \iiint_{-\infty}^{\infty} G(\mathbf{x}_{\mathrm{c}})\, \mathrm{e}^{\mathrm{i}\mathbf{k}_{\mathrm{c}}^{\mathrm{T}}\mathbf{x}_{\mathrm{c}}}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z. \tag{2.32}$$

The inverse transform is given by

$$G(\mathbf{x}_{\mathrm{c}}) = \tilde{\mathcal{F}}_{\mathbf{x}}^{-1}\{G(\mathbf{k}_{\mathrm{c}})\} = \frac{1}{(2\pi)^3}\iiint_{-\infty}^{\infty} \tilde{G}(\mathbf{k}_{\mathrm{c}})\, \mathrm{e}^{-\mathrm{i}\mathbf{k}_{\mathrm{c}}^{\mathrm{T}}\mathbf{x}_{\mathrm{c}}}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z. \tag{2.33}$$

Finally, the time-frequency Fourier transform and the spatial Fourier transform can also be joined by the same approach of superposition describing a four-dimensional signal yielding

$$\tilde{G}(\mathbf{k}_{\mathrm{c}}, \omega) = \tilde{\mathcal{F}}_{\mathbf{x}\mathrm{t}}\{g(\mathbf{x}_{\mathrm{c}}, t)\} = \iiiint_{-\infty}^{\infty} g(\mathbf{x}_{\mathrm{c}}, t)\, \mathrm{e}^{i(\mathbf{k}_{\mathrm{c}}^{\mathrm{T}}\mathbf{x}_{\mathrm{c}} - \omega\, t)}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}t. \tag{2.34}$$

The inverse transform is given by

$$g(\mathbf{x}_\mathrm{c}, t) = \tilde{\mathcal{F}}_{\mathbf{x}\mathrm{t}}^{-1}\{G(\mathbf{k}_\mathrm{c}, \omega)\} = \frac{1}{(2\pi)^4} \iiiint_{-\infty}^{\infty} \tilde{G}(\mathbf{k}_\mathrm{c}, \omega) \mathrm{e}^{-\mathrm{i}(\mathbf{k}_\mathrm{c}^\mathrm{T}\mathbf{x}_\mathrm{c} - \omega\, t)}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}t. \tag{2.35}$$

In an analogous manner, the multidimensional transform relations can be extended to arbitrary higher-dimensional signals and spaces $\mathbb{R}^\nu, \nu \in \mathbb{N}$ as proposed in (Spors, 2006, pp 44–45). However, in this thesis all problems are covered using transforms with four or less dimensions. The superimposed transform relations Eq. (2.34) and Eq. (2.35) show the separability of the Fourier transforms.

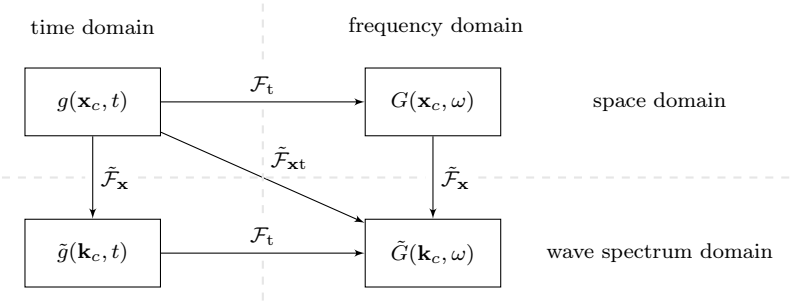


Figure 2.2 Transform relations and respective domains for the temporal and the spatial Fourier transforms in Cartesian coordinates. The diagram shows the forward transforms. In the corresponding scheme for the inverse transforms the arrows are reversed. The diagram illustrates the separability of the temporal and spatial transforms (Spors, 2006, p 45).

Instead of applying the entire transform operation at once, temporal and spatial Fourier transforms can be applied sequentially (Spors, 2006, pp 44–45) yielding

$$\tilde{G}(\mathbf{k}_\mathrm{c}, \omega) = \tilde{\mathcal{F}}_\mathbf{x}\Big\{\mathcal{F}_\mathrm{t}\big\{g(\mathbf{x}, t)\big\}\Big\} = \mathcal{F}_\mathrm{t}\Big\{\tilde{\mathcal{F}}_\mathbf{x}\big\{g(\mathbf{x}, t)\big\}\Big\} = \tilde{\mathcal{F}}_\mathbf{xt}\big\{g(\mathbf{x}, t)\big\}, \tag{2.36}$$

which for the inverse transforms reads

$$g(\mathbf{x}, t) = \mathcal{F}_\mathrm{t}^{-1}\Big\{\tilde{\mathcal{F}}_\mathbf{x}^{-1}\big\{\tilde{G}(\mathbf{k}_\mathrm{c}, \omega)\big\}\Big\} = \tilde{\mathcal{F}}_\mathbf{x}^{-1}\Big\{\mathcal{F}_\mathrm{t}^{-1}\big\{\tilde{G}(\mathbf{k}_\mathrm{c}, \omega)\big\}\Big\} = \tilde{\mathcal{F}}_\mathbf{xt}^{-1}\big\{\tilde{G}(\mathbf{k}_\mathrm{c}, \omega)\big\}. \tag{2.37}$$

Figure 2.2 gives an overview of the transform relations in Cartesian coordinates and the respective Fourier domains that are of importance for this thesis.

2.5 Helmholtz Equation

A different formulation of the wave equation is given by the Helmholtz equation that delivers stationary solutions in the frequency domain. The Helmholtz equation and its solutions are of fundamental importance for this thesis. The Helmholtz equation can be obtained by applying a time-frequency Fourier transform to Eq. (2.1), (Williams, 1999, p 18). The time-frequency Fourier transform is given in Eq. (2.28) and the inverse transform in Eq. (2.29). The function prototype $G(\omega)$ in Eq. (2.29) is replaced by the spatial pressure spectrum $P(\mathbf{x}, \omega)$ and $g(t)$ by the spatial pressure function $p(\mathbf{x}, t)$ therein. Next, Eq. (2.29) is differentiated with respect to the time yielding

$$\frac{\partial}{\partial t}\, p(\mathbf{x}, t) = \frac{\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} \omega\, P(\mathbf{x}, \omega)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}\omega. \tag{2.38}$$

A second derivation with respect to time gives

$$\frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2\, P(\mathbf{x}, \omega)\, \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}\omega. \tag{2.39}$$

The following correspondences arise:

$$\frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) = -\mathcal{F}_{\mathrm{t}}^{-1}\big\{\omega^2\, P(\mathbf{x}, \omega)\big\}, \tag{2.40}$$

and

$$\mathcal{F}_{\mathrm{t}}\left\{ \frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) \right\} = -\omega^2 P(\mathbf{x}, \omega). \tag{2.41}$$

In a next step, the time-frequency Fourier transform from Eq. (2.28) is applied to the wave equation Eq. (2.1) giving

$$\mathcal{F}_{\mathrm{t}}\left\{ \nabla^2 p(\mathbf{x}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) \right\} = 0 \tag{2.42}$$

$$\Leftrightarrow \mathcal{F}_{\mathrm{t}}\Big\{ \nabla^2 p(\mathbf{x}, t) \Big\} - \frac{1}{c^2}\, \mathcal{F}_{\mathrm{t}}\left\{ \frac{\partial^2}{\partial t^2}\, p(\mathbf{x}, t) \right\} = 0, \tag{2.43}$$

The first term does not have an operator with respect to the time t. As a consequence, it can be transformed by direct application of the Fourier transform to $p(\mathbf{x}, t)$ yielding

$$\nabla^2 P(\mathbf{x}, \omega) - \frac{1}{c^2} \mathcal{F}_\mathrm{t} \left\{ \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) \right\} = 0. \tag{2.44}$$

The transform of the second term yields

$$\nabla^2 P(\mathbf{x}, \omega) - \frac{1}{c^2} \left(-\omega^2\right) P(\mathbf{x}, \omega) = 0, \tag{2.45}$$

which finally leads to the well-known Helmholtz equation that reads

$$\nabla^2 P(\mathbf{x}, \omega) + \left(\tfrac{\omega}{c}\right)^2 P(\mathbf{x}, \omega) = 0. \tag{2.46}$$

The quotient ω/c is often replaced by the wave number k according to the dispersion relation of Eq. (2.10).

2.6 Solutions to the Helmholtz Equation in Cartesian Coordinates

Analogous to obtaining the Helmholtz equation by applying the time-frequency Fourier transform to the wave equation from Eq. (2.1), a solution to the Helmholtz equation can be found by applying the time-frequency Fourier transform to Eq. (2.12) yielding

$$P(\mathbf{x}_\mathrm{c}, \omega) = \widehat{P}\, \mathrm{e}^{-\mathrm{i}\, \mathbf{k}_c^\mathrm{T} \mathbf{x}_\mathrm{c}} = \mathcal{F}_\mathrm{t} \Big\{ \widehat{p}\, \mathrm{e}^{-\mathrm{i}(\mathbf{k}_c^\mathrm{T} \mathbf{x}_\mathrm{c} - \omega t)} \Big\}. \tag{2.47}$$

It is conspicuous that the ωt-term in the exponential disappears. To study this aspect, a dedicated frequency $\omega = \omega_0$ is inserted into the initial equation and the exponential function is split according to

$$\widehat{p}\, \mathrm{e}^{\mathrm{i}\, \omega_0 t}\, \mathrm{e}^{-\mathrm{i}\, \mathbf{k}_{c,0}^\mathrm{T} \mathbf{x}_\mathrm{c}} = \widehat{p}\, \mathrm{e}^{-\mathrm{i}(\mathbf{k}_{c,0}^\mathrm{T} \mathbf{x}_\mathrm{c} - \omega_0 t)}. \tag{2.48}$$

Consequently, the wave vector that implicitly involves ω must also be reformulated with respect to ω_0. This is denoted by attaching 0 to the wave vector's index. Taking the frequency-shift property of the Fourier transform (Beerends et al., 2003) into account, the following correspondence arises:

$$\mathcal{F}_\mathrm{t} \Big\{ \mathrm{e}^{\mathrm{i}\omega_0 t} \Big\} = 2\pi\, \delta(\omega - \omega_0), \tag{2.49}$$

where δ denotes the Dirac delta function (Arfken and Weber, 2005, pp 83–95). Thus the time-frequency Fourier transform of the expression given in Eq. (2.48) yields

$$P'(\mathbf{x}_\mathrm{c}, \omega) = \widehat{p}\, 2\pi\, \delta(\omega - \omega_0)\, \mathrm{e}^{-\mathrm{i}\, \mathbf{k}_{c,0}^\mathrm{T} \mathbf{x}_\mathrm{c}} = \mathcal{F}_\mathrm{t}\Big\{ \widehat{p}\, \mathrm{e}^{-\mathrm{i}(\mathbf{k}_{c,0}^\mathrm{T} \mathbf{x}_\mathrm{c} - \omega_0 t)} \Big\}. \tag{2.50}$$

This expression points out the monochromatic nature of the time dependence (Williams, 1999, p 21). The sifting-property of the Dirac delta function (Beerends et al., 2003, p 193) generally sets all the expression to zero except for the single specific frequency ω_0, sifting out one particular valid solution. To come back to a more elegant expression again, the selected frequency ω_0 is reset to ω and hence the argument of the δ-function keeps constantly zero, bypassing the δ-function as it always exactly sifts out the currently observed monochromatic frequency. Thus, it can be omitted and set to one. By defining a constant $\widehat{P} = 2\pi\, \widehat{p}$ the expression $P'(\mathbf{x}_\mathrm{c}, \omega)$ becomes $P(\mathbf{x}_\mathrm{c}, \omega)$ given in the initial equation Eq. (2.47).

2.7 Plane Wave Expansion in Cartesian Coordinates

The previous considerations referred to single plane waves. Complex source-free sound fields can be expressed by an appropriately weighted summation of single plane waves (Williams, 1999, pp 31–32), (Spors, 2006, pp 13–14), which is referred to as plane wave expansion. This is analogous to the classic Fourier series where complex time-domain signals are described by the sum of suitably weighted sinoid signals (Beerends et al., 2003, pp 60–133). Let $S(\mathbf{x}_\mathrm{c}, \omega)$ denote a complex sound field at a specific angular frequency ω characterized by a stationary pressure distribution in Cartesian coordinates. Then $S(\mathbf{x}_\mathrm{c}, \omega)$ is expected to be expressible in terms of plane waves according to the plane wave solution given in Eq. (2.47). Specific coefficients $\tilde{P}$ are introduced in order to weight the plane waves with respect to their amplitude and phase. These coefficients are called expansion coefficients. As previously shown, the three components of the wave vector are not independent, compare Eq. (2.17). Thus, if the values for two of the wave vectors components are chosen, the value of the third one is determined except for its sign; this is reflected in Eq. (2.19) to Eq. (2.24). Let k_z be the dependent wave vector component. Consequently, only value pairs for k_x and k_y are regarded for the summation, since the value for k_z emerges according to Eq. (2.21) and Eq. (2.24). Without any loss of generality may the dependent value range for k_z be restricted to the positive branch of the respective square roots. Thus, all sources are located in the negative z-half space and waves do only travel in positive z-direction. The values for

$\{k_\mathrm{x}, k_\mathrm{y}\} \in \mathbb{R}$ extend over the entire range from $-\infty$ to ∞. After that, a sound field $S(\mathbf{x}_\mathrm{c}, \omega)$ can be expressed as

$$S(\mathbf{x}_\mathrm{c}, \omega) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega)\, \mathrm{e}^{-\mathrm{i}\, \mathbf{k}_c^\mathrm{T} \mathbf{x}_\mathrm{c}} \mathrm{d}k_\mathrm{x}\, \mathrm{d}k_\mathrm{y}. \tag{2.51}$$

The dependence on the angular frequency is implicitly given through the wave vector. The amplitude scaling factor $1/(4\pi^2)$ is introduced in order to close the connection to the spatial Fourier transform, which will become explicit in the following. The expansion equation given in Eq. (2.51) is quite useful to understand the expansion principle, however, its true potential does not reveal until the inverse operation and the subsequent operations, i.e. sound field extrapolation or plane wave decomposition, are discussed. It is best to derive the inverse operation first. An inverse operation to Eq. (2.51) would extract expansion coefficients $\tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega)$ from a complex sound field $S(\mathbf{x}_\mathrm{c}, \omega)$. Without any loss of generality z can be set to zero and reducing our observation to the x-y-plane Eq. (2.51) yields

$$S(\mathbf{x}_\mathrm{c}|_{z=0}, \omega) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega)\, \mathrm{e}^{-\mathrm{i}\,(k_\mathrm{x} x + k_\mathrm{y} y)}\, \mathrm{d}k_\mathrm{x}\, \mathrm{d}k_\mathrm{y}, \tag{2.52}$$

rewritten in a slightly different manner as

$$S(\mathbf{x}_\mathrm{c}|_{z=0}, \omega) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega)\, \mathrm{e}^{-\mathrm{i}\, k_\mathrm{x} x}\, \mathrm{e}^{-i\, k_\mathrm{y} y}\, \mathrm{d}k_\mathrm{x}\, \mathrm{d}k_\mathrm{y}, \tag{2.53}$$

and compared to Eq. (2.31) on page 33, the expression turns out to be an inverse two-dimensional spatial Fourier transform. This obviously simplifies the inversion of the expression. Analogous to the spatial Fourier transform Eq. (2.30) on page 33 an expression for gaining expansion coefficients can be given by

$$\tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega) = \iint_{-\infty}^{\infty} S(\mathbf{x}_\mathrm{c}|_{z=0}, \omega)\, \mathrm{e}^{\mathrm{i}\,(k_\mathrm{x} x + k_\mathrm{y} y)}\, \mathrm{d}x\, \mathrm{d}y\,. \tag{2.54}$$

The expansion coefficients describing the sound field in the wave spectrum domain can be used for sound field extrapolation and plane wave decomposition. The plane wave decomposition transforms the wave field into plane wave components referring to a specific observation angle. Due to the dependency on an observation angle, cylindrical coordinates (Spors, 2006, pp 56–100) or spherical coordinates, cf. Section 2.9 are better suited to this problem than Cartesian coordinates. The plane wave decomposition is therefore discussed later after the introduction of the spherical coordinate system and

its inherent sound field equations. The sound field extrapolation, on the other hand, is performed most comprehensively in the Cartesian coordinate system.

2.8 Sound Field Extrapolation

Sound field extrapolation is based on the Fourier expansion of a sound field. The advantage of this technique is that a source free sound field can be constructed based on the knowledge of the field parameters on a single plane or cylindrical or spherical surface only. In the Cartesian coordinate system this technique is easy to grasp. As already shown in Eq. (2.18) et seq., the three wave vector components are not fully independent. This has a direct impact on the equations describing the plane wave expansion, Eq. (2.51) to Eq. (2.54), where k_z is a dependent variable and the Fourier integrals depend on k_x and k_y only. In a first step, the expansion coefficients $\tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega)$ for the x-y plane are calculated by setting the z-coordinate to zero and using Eq. (2.54), which obviously requires knowledge of the sound field in the respective x-y plane only. The sound field in a different parallel plane with $z' > 0$ can be calculated with the dependency of k_z given in the relations Eq. (2.16) et seq. and using Eq. (2.51):

$$S(\mathbf{x}'_\mathrm{c}|_{z=z'}, \omega) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \tilde{P}(k_\mathrm{x}, k_\mathrm{y}, \omega) \mathrm{e}^{-\mathrm{i}\left(k_\mathrm{x}x + k_\mathrm{y}y + \sqrt{(\omega/c)^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2}\, z'\right)} \,\mathrm{d}k_\mathrm{x}\,\mathrm{d}k_\mathrm{y}. \tag{2.55}$$

The entire sound field for the positive half space $z > 0$ can be calculated based on the knowledge of the sound field in the x-y plane at $z = 0$ only. This extrapolation, as well as an alternative point of view, are discussed in (Williams, 1999, p 31-33). A derivation of the Rayleigh integrals based on the extrapolation approach is given in (Williams, 1999, pp 34–37). Extrapolation techniques can also be adapted to different scenarios and coordinate systems. Extrapolation in a spherical coordinate system will be discussed later.

2.9 Helmholtz Equation in Spherical Coordinates

The sound field equations can be formulated for different coordinate systems. The choice of the appropriate coordinate system depends on the respective problem. Common coordinate systems, besides the classical Cartesian coordinates, are the cylindrical and the spherical coordinate systems. The latter is discussed in the following. For a detailed discussion of the sound field equations in cylindrical coordinates the reader may refer to (Williams, 1999, pp 115–181) or (Spors, 2006, e.g. pp 14-18).

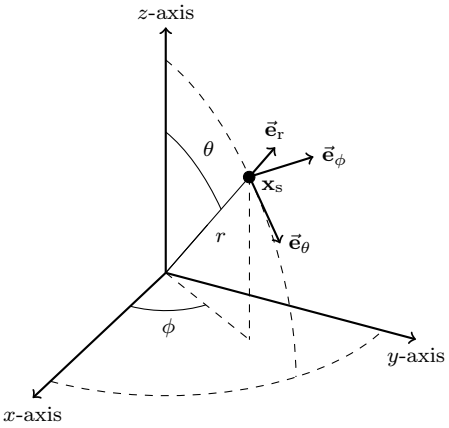


Figure 2.3 Illustration of the $\mathbf{x}_\mathrm{s}$ vector in a spherical coordinate system as used in this thesis; $\vec{\mathbf{e}}_\theta$, $\vec{\mathbf{e}}_\phi$ and $\vec{\mathbf{e}}_\mathrm{r}$ indicate the corresponding unit vectors.

The Helmholtz equation Eq. (2.46) and its stationary solutions are specified in a spherical coordinate system depicted in Figure 2.3. This configuration is of major interest for this thesis. The time domain wave equation Eq. (2.1) and its solutions can be expressed analogous.

The position vector in spherical coordinates is denoted as

$$\mathbf{x}_\mathrm{s} = r\,\vec{\mathbf{e}}_\mathrm{r} + \theta\,\vec{\mathbf{e}}_\theta + \phi\,\vec{\mathbf{e}}_\phi, \tag{2.56}$$

where $\vec{\mathbf{e}}_\theta$, $\vec{\mathbf{e}}_\phi$, and $\vec{\mathbf{e}}_\mathrm{r}$ denote the corresponding unit vectors. The relation between the Cartesian unit vectors $\vec{\mathbf{e}}_\mathrm{x}$, $\vec{\mathbf{e}}_\mathrm{y}$, and $\vec{\mathbf{e}}_\mathrm{z}$ is given by a transform matrix

$$\begin{bmatrix} \vec{\mathbf{e}}_\mathrm{r} \\ \vec{\mathbf{e}}_\theta \\ \vec{\mathbf{e}}_\phi \end{bmatrix} = \begin{vmatrix} \sin\theta\,\cos\phi & \sin\theta\,\sin\phi & \cos\theta \\ \cos\theta\,\cos\phi & \cos\theta\,\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{vmatrix} \begin{bmatrix} \vec{\mathbf{e}}_\mathrm{x} \\ \vec{\mathbf{e}}_\mathrm{y} \\ \vec{\mathbf{e}}_\mathrm{z} \end{bmatrix}. \tag{2.57}$$

Since the transform matrix is orthogonal, the inverse operation can be obtained by a matrix transposition $|\cdot|^\mathbf{T}$ of the transform matrix given in Eq. (2.57).

The gradient of a function $f(r, \theta, \phi)$ is given by

$$\nabla_\mathrm{s} f = \frac{\partial f}{\partial r}\vec{\mathbf{e}}_\mathrm{r} + \frac{1}{r}\,\frac{\partial f}{\partial \theta}\vec{\mathbf{e}}_\theta + \frac{1}{r\,\sin\,\theta}\,\frac{\partial f}{\partial \phi}\vec{\mathbf{e}}_\phi. \tag{2.58}$$

The divergence of a vector field $\vec{\mathbf{f}}$ is defined as

$$\nabla_{\mathrm{c}} \cdot \vec{\mathbf{f}} = \frac{1}{r^2} \frac{\partial (r^2 \, f_{\mathrm{r}})}{\partial r} + \frac{1}{r \, \sin \, \theta} \frac{\partial}{\partial \theta} (f_\theta \, \sin \, \theta) + \frac{1}{r \, \sin \, \theta} \frac{\partial f_\phi}{\partial \phi}, \tag{2.59}$$

with

$$\vec{\mathbf{f}} = \begin{bmatrix} f_{\mathrm{r}}(r, \theta, \phi) \\ f_\theta(r, \theta, \phi) \\ f_\phi(r, \theta, \phi) \end{bmatrix}. \tag{2.60}$$

The Laplacian is defined as divergence of the gradient of a function $f(x, y, z)$ in Cartesian coordinates,

$$\nabla_{\mathrm{s}}^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \Big(r^2 \frac{\partial f}{\partial r} \Big) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \Big(\sin \theta \, \frac{\partial P}{\partial \theta} \Big) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}. \tag{2.61}$$

Thus, the Helmholtz equation in spherical coordinates explicitly reads

$$\frac{1}{r^2} \frac{\partial}{\partial r} \Big(r^2 \frac{\partial P}{\partial r} \Big) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \Big(\sin \theta \, \frac{\partial P}{\partial \theta} \Big) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 P}{\partial \phi^2} + \Big(\frac{w}{c} \Big)^2 P = 0 \, . \tag{2.62}$$

2.10 Solutions to the Helmholtz Equation in Spherical Coordinates

The Helmholtz equation in spherical coordinates, cf. Eq. (2.62), can be solved by a separation of variables. The expression is decomposed into different equations, where each depends on a single variable only. For the separated differential equations explicit solutions are known. The separation of variables and the solutions to the separated equations are discussed according to Williams (1999), Blackstock (2000), and Jin (2011) in the following.

2.10.1 Separation of Variables

We assume that Eq. (2.62) is separable and that the solutions to the equation can be expressed in product form by using expressions $R(r)$, $\Theta(\theta)$, and $\Phi(\phi)$ that represent solutions to the separated equations. This yields

$$P(r, \theta, \phi) = R(r) \cdot \Theta(\theta) \cdot \Phi(\phi). \tag{2.63}$$

This expression is substituted in Eq. (2.62). The full equation is divided by $R(r)\,\Theta(\theta)\,\Phi(\phi)$ and multiplied by $(r^2 \sin^2\theta)$ yielding

$$\frac{\sin^2\theta}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\frac{\sin\theta}{\Theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)+\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2}+\left(\frac{\omega}{c}\right)^2 r^2\sin^2\theta=0. \tag{2.64}$$

Since only the third term depends on ϕ, the first separation can be directly deduced (Jin, 2011, p 249) giving

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2}+m^2\Phi=0, \tag{2.65}$$

where m^2 denotes a constant that depends on the specific problem. Thus Eq. (2.64) is reduced to

$$\frac{\sin^2\theta}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\frac{\sin\theta}{\Theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)+\left(\frac{\omega}{c}\right)^2 r^2\sin^2\theta-m^2=0. \tag{2.66}$$

Dividing by $(\sin^2\theta)$ and rearranging according to terms depending on r and terms depending on θ yields

$$\left[\frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\left(\frac{\omega}{c}\right)^2 r^2\right]+\left[\frac{1}{\Theta\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)-\frac{m^2}{\sin^2\theta}\right]=0. \tag{2.67}$$

Eq. (2.67) can be separated into two expressions (Jin, 2011, p 249):

$$\frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\left(\frac{\omega}{c}\right)^2 r^2=n(n+1),\ \text{and} \tag{2.68}$$

$$\frac{1}{\Theta\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)-\frac{m^2}{\sin^2\theta}=-n(n+1). \tag{2.69}$$

The term $n(n+1)$ denotes a constant depending on the specific problem. After slightly rearranging Eq. (2.68) and Eq. (2.69) and picking up Eq. (2.65) again, the Helmholtz equation is finally decomposed into three ordinary differential equations (Williams, 1999, p 185):

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+\left[\left(\frac{\omega}{c}\right)^2 r^2-n(n+1)\right]R=0, \tag{2.70}$$

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)+\left[n(n+1)-\frac{m^2}{\sin^2\theta}\right]\Theta=0, \text{ and} \tag{2.71}$$

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2}+m^2\Phi=0. \tag{2.65}$$

Each equation depends either on r, θ, or ϕ only. All three equations and their solutions are well-known. Applying minor changes and substitutions (Williams, 1999, pp 193–194), Eq. (2.70) can be transformed into a spherical Bessel differential equation (Abramowitz and Stegun, 1972, p 437). Eq. (2.71) is known as Legendre differential equation (Abramowitz and Stegun, 1972, p 332) and Eq. (2.65) is an ordinary second-order differential equation.

2.10.2 Spherical Bessel and Hankel Functions

The spherical Bessel equation, cf. Eq. (2.69), that represents the radial component of the Helmholtz equation is solved by nth-order spherical Bessel functions of the first kind j_n and of the second kind y_n, and spherical Hankel functions of the first kind $h_n^{(1)}$ and of the second kind $h_n^{(2)}$ (Abramowitz and Stegun, 1972, p 437). The spherical Bessel functions of the second kind are also referred to as spherical Neumann functions and the spherical Hankel functions as spherical Bessel functions of the third kind. The spherical Bessel functions of the first and the second kind for integer orders are depicted in Figure 2.4. The spherical Hankel functions are complex-valued linear combinations of spherical Bessel functions of the first and the second kind. This is similar to the exponential function, which represents a complex composition of trigonometric functions that is well-known as Euler's formula. The spherical Hankel functions are defined as (Abramowitz and Stegun, 1972, p 437)

$$h_n^{(1)}(z)=j_n(z)+\mathrm{i}\,y_n(z), \text{ and} \tag{2.72}$$

$$h_n^{(2)}(z)=j_n(z)-\mathrm{i}\,y_n(z). \tag{2.73}$$

Whenever the number indicating the kind of Hankel function is omitted, the spherical Hankel function of the second kind is addressed throughout this work, i.e. $h_n=h_n^{(2)}$. Trigonometric expressions describing the spherical Bessel and Hankel functions, Maclaurin/Taylor series, the relation to the conventional Bessel and Hankel functions, as well

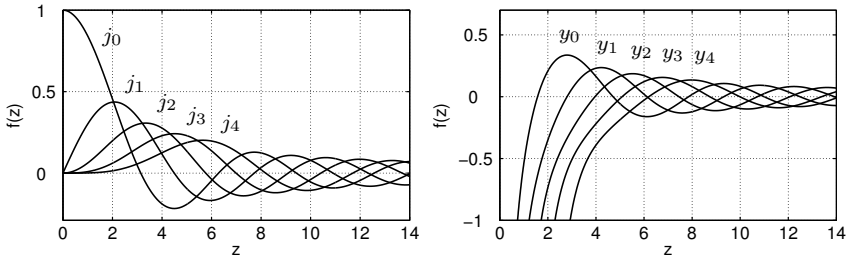


Figure 2.4 Spherical Bessel functions of the first kind $j_n(z)$ and of the second kind $y_n(z)$ each for integer orders $n = \{0, 1, 2, 3, 4\}$.

as several mathematical relations concerning spherical Bessel and Hankel functions are discussed in (Abramowitz and Stegun, 1972, pp 437–442), (Williams, 1999, pp 193–197), and (Blackstock, 2000, pp 341–345). A general solution $R(r)$ to Eq. (2.70) is given by a linear combination of spherical Bessel functions of the first and of the second kind (Williams, 1999, p 185):

$$R(r) = \psi_1 \, j_n\left(\tfrac{\omega}{c} \, r\right) + \psi_2 \, y_n\left(\tfrac{\omega}{c} \, r\right). \tag{2.74}$$

Solutions can also be expressed as (Williams, 1999, p 186)

$$R(r) = \psi_3 \, h_n^{(1)}\left(\tfrac{\omega}{c} \, r\right) + \psi_4 \, h_n^{(2)}\left(\tfrac{\omega}{c} \, r\right), \tag{2.75}$$

where ψ_1 to ψ_4 denote arbitrary constants. The specific choice of the spherical Bessel or Hankel functions depends on the observed scenario and the location of the sources. More details on this topic are discussed in subsequent sections.

2.10.3 Legendre Polynomials and Associated Legendre Functions

The expression that depends on the elevation angle θ, given in Eq. (2.71), is solved by Legendre polynomials $P_n(\cos\theta)$ or their generalization given by associated Legendre functions of the first kind, $P_n^m(\cos\theta)$, (Jin, 2011, pp 249–250), (Abramowitz and Stegun, 1972, p 332), (Williams, 1999, p 185); n is referred to as order and m as mode of the respective functions or polynomials. The cosine function in the argument arises from a transform of variables that is used to transform a given equation to the native Legendre differential equation, i.e. to map the elevation angle range from $[0, \pi]$ to the associated Legendre function's domain from $[-1, 1]$. Legendre polynomials can be represented

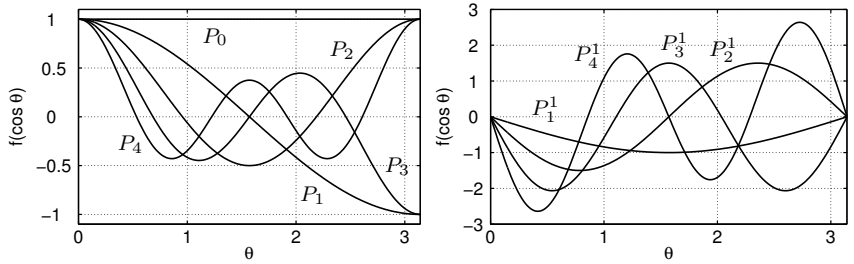


Figure 2.5 Legendre polynomials $P_n(\cos\theta)$ for $n = \{0, 1, 2, 3, 4\}$ and exemplary associated Legendre functions $P_n^m(\cos\theta)$ for $m = 1$ and $n = \{1, 2, 3, 4\}$. The angle θ is defined in the range $\theta \in [0, \pi]$.

in a compact expression referred to as Rodrigues' formula (Atkinson and Han, 2012, pp 36–39), (Jackson, 1962, p 57) that reads

$$P_n(z) = \frac{1}{2^n n!} \frac{d^n}{dz^n}(z^2 - 1)^n. \tag{2.76}$$

For $m = 0$ the associated Legendre functions $P_n^m(x)$ are identical to the Legendre polynomials $P_n(x)$:

$$P_n^m(z)|_{m=0} = P_n^0(z) = P_n(z). \tag{2.77}$$

For $m > 0$ the relation between both is given by (Jackson, 1962, p 64)

$$P_n^m(z) = (-1)^m (1 - z^2)^{m/2} \frac{d^m}{dz^m} P_n(z), \tag{2.78}$$

and for $m < 0$ the following equation can be applied (Jackson, 1962, p 65):

$$P_n^{-m}(z) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(z). \tag{2.79}$$

The $(-1)^m$ factor is a sign convention referred to as Condon-Shortley phase (Arfken and Weber, 2005, p 788) after Condon and Shortley (1951); it was originally introduced in the context of quantum mechanics in order to simplify the treatment of angular momentum. Using Rodrigues' formula (Atkinson and Han, 2012, pp 36–39) to represent

the Legendre polynomial in Eq. (2.78), a common expression for both positive and negative values for m can be found that yields (Jackson, 1962, p 64)

$$P_n^m(z) = \frac{(-1)^m}{2^n n!}(1-z^2)^{m/2}\,\frac{d^{n+m}}{dz^{n+m}}(z^2-1)^n. \tag{2.80}$$

Legendre polynomials and associated Legendre functions are orthogonal function sets. This property is of fundamental importance and discussed in the context of spherical harmonics later in Section 2.10.5. General solutions to the Legendre differential equation may involve Legendre functions of the second kind $Q_n^m(\cos\theta)$ (Abramowitz and Stegun, 1972, p 332). Since the latter are not finite at the poles, $\theta = \{0, \pi\}$, the respective solutions are discarded and the solutions to Eq. (2.71) are given by (Williams, 1999, p 185)

$$\Theta(\theta) = \chi_1\, P_n^m(\cos\theta) + \chi_2\, Q_n^m(\cos\theta), \tag{2.81}$$

where χ_1 and χ_2 are arbitrary constants. χ_2 is generally set to zero in order to explicitly exclude any Legendre functions of the second kind from the solutions. Trigonometric expressions describing the Legendre polynomials and associated Legendre functions, respective series expansions, as well as several further mathematical relations are discussed in (Abramowitz and Stegun, 1972, pp 332–339), (Williams, 1999, pp 186–191), and (Blackstock, 2000, pp 338–341).

2.10.4 Solutions to the Separated Azimuthal Equation

The separated azimuthal equation, cf. Eq. (2.65), which depends on the azimuthal angle $\phi \in [0, 2\pi]$ is an ordinary second-order differential equation that can be solved by harmonic functions, such as complex exponential functions, yielding (Williams, 1999, p 185)

$$\Phi(\phi) = \zeta_1\, \mathrm{e}^{im\phi} + \zeta_2\, \mathrm{e}^{-im\phi}, \tag{2.82}$$

where ζ_1 and ζ_2 denote arbitrary constants.

2.10.5 Spherical Harmonics

The angular portion, consisting of both angular functions for elevation θ and azimuth ϕ, can conveniently be described by a single function $Y_n^m(\theta, \phi)$ that is referred to as spherical harmonics or surface spherical harmonics and defined by (Jackson, 1962, p 65)

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\,\frac{(n-m)!}{(n+m)!}}\, P_n^m(\cos\theta)\, \mathrm{e}^{im\phi}. \tag{2.83}$$

According to Eq. (2.79) negative values for m yield (Jackson, 1962, p 65)

$$Y_n^{-m}(\theta,\phi) = (-1)^m\, Y_n^m(\theta,\phi)^*. \tag{2.84}$$

The spherical harmonics given in Eq. (2.83) are composed of associated Legendre functions and complex exponential functions. A normalization factor is introduced that is different for different scientific disciplines. The normalization used here entails orthonormality of the spherical harmonics and is commonly used in classical physics. Orthonormality over a spherical surface yields (Jackson, 1962, p 65), (Arfken and Weber, 2005, p 788)

$$\int_0^{2\pi}\int_0^{\pi} Y_n^m(\theta,\phi)\; Y_{n'}^{m'}(\theta,\phi)^*\; \mathrm{d}\phi\, \sin\theta\, \mathrm{d}\theta = \delta_{nn'}\, \delta_{mm'}, \tag{2.85}$$

with

$$\delta_{ll'} = \begin{cases} 1 & \text{if } l = l' \\ 0 & \text{if } l \neq l' \end{cases} \tag{2.86}$$

denoting the Kronecker delta (Arfken and Weber, 2005, p 10). Atkinson and Han (2012) show in great detail that spherical harmonics are the only irreducible system of function spaces that is complete and closed. The completeness relation (Jackson, 1962, pp 44–47) in spherical harmonics yields (Jackson, 1962, p 65)

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^m(\theta,\phi) Y_n^m(\theta',\phi')^* = \delta(\phi-\phi')\,\delta(\cos\theta - \cos\theta'), \tag{2.87}$$

where δ now denotes a Dirac delta function (Arfken and Weber, 2005, pp 83–95). Trigonometric expressions describing spherical harmonics, simplifications for special problems, as well as several other mathematical relations are discussed in (Atkinson and Han, 2012, pp 11–81), (Jackson, 1962, pp 66–69), (Williams, 1999, pp 192–193), (Arfken and Weber, 2005, p 790), and (Varshalovich et al., 1988, pp 130–163).

The real and imaginary parts and the magnitudes of the first spherical harmonics for orders $n = \{0...4\}$ including all possible corresponding modes m are shown in Figure 2.6 and Figure 2.7.

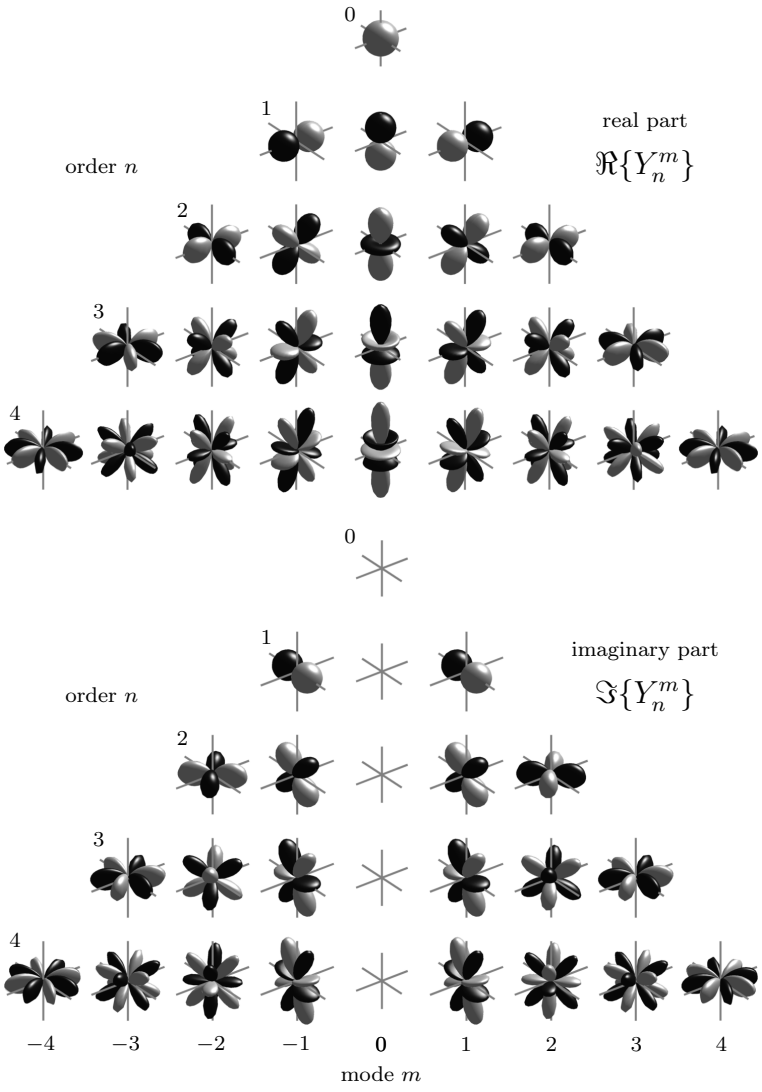


Figure 2.6 Real $\Re\{Y_n^m\}$ part and imaginary $\Im\{Y_n^m\}$ part of the spherical harmonics. The polarity is indicated by gray (plus) and black (minus) colors. The $m = 0$ responses of $\Re\{Y_n^m\}$ are diminished by 3 dB for better illustration.

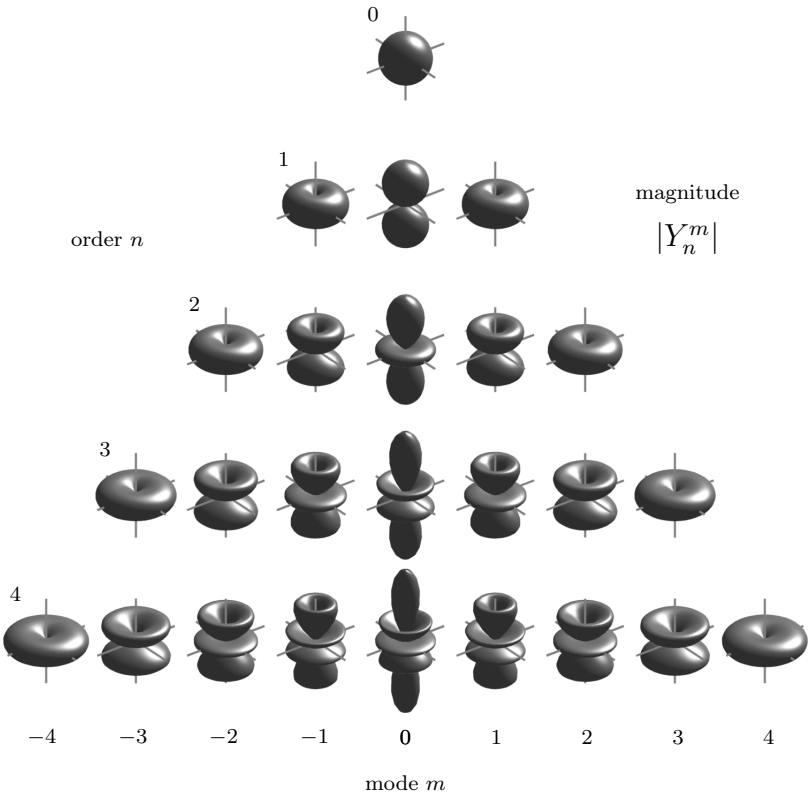


Figure 2.7 Directional magnitudes of the spherical harmonics $|Y_n^m|$.

2.11 Spherical Harmonic Expansion of Functions on the Sphere

In addition to solving the angular portion of the Helmholtz equation, spherical harmonics can be used to expand arbitrary functions $G(\theta, \phi)$ on the (unit) sphere (Williams, 1999, p 192):

$$G(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}_{nm} \, Y_n^m(\theta, \phi), \tag{2.88}$$

with complex constants $\mathring{G}_{nm}$ that are referred to as expansion coefficients. This operation reveals the true power of the spherical harmonics as will become explicit during the following sections. The rough structure of the expanded function $G(\theta, \phi)$ is modeled by the lower order spherical harmonics and the fine structure by higher order spherical harmonics. This observation is important in case the expansion is truncated at a finite order N, which is discussed in subsequent chapters. Due to the orthonormality of the spherical harmonics, cf. Eq. (2.85), the complex coefficients $\mathring{G}_{nm}$ can be found using (Williams, 1999, p 192)

$$\mathring{G}_{nm} = \int_0^{2\pi} \int_0^{\pi} G(\theta, \phi) \, Y_n^m(\theta, \phi)^* \, \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi. \tag{2.89}$$

Hence, the coefficients $\mathring{G}_{nm}$ are calculated by integrating the function $G(\theta, \phi)$ over the full (unit) sphere. $G(\theta, \phi)$ is weighted with the complex conjugate spherical harmonic function of the respective order n and mode m.

2.12 Spatial Fourier Transform in Spherical Coodinates

In a next step, the function $G(\theta, \phi)$ on the unit sphere is replaced by a more specific function $G(\theta, \phi, r_0, \omega)$ that is defined on a radius r_0 for a specific angular frequency ω. This function represents a sound pressure distribution $P(\theta, \phi, r_0, \omega)$, for instance. Inserting $G(\theta, \phi, r_0, \omega)$ into equation Eq. (2.89) yields

$$\mathring{G}_{nm}(r_0, \omega) = \int_0^{2\pi} \int_0^{\pi} G(\theta, \phi, r_0, \omega) \, Y_n^m(\theta, \phi)^* \, \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi, \tag{2.90}$$

which is referred to as forward spatial Fourier transform in spherical coordinates. In order to simplify notation and handling, the solid angle Ω over the spherical surface S is introduced (Arfken and Weber, 2005, p 124) with $\mathrm{d}\Omega = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$. We use

$$\iint_S \mathrm{d}\Omega = \int_0^{2\pi} \int_0^{\pi} \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi \tag{2.91}$$

to represent the integration over the full sphere S. Let now be S_0 a sphere of radius r_0. The position angles θ and ϕ, as well as the radius r_0 can be combined in a single vector $\mathbf{x}_{\mathrm{s},0} \in S_0$ according to Eq. (2.56) yielding

$$\mathring{G}_{nm}(r_0, \omega) = \iint_{S_0} G(\mathbf{x}_{\mathrm{s},0}, \omega)\, Y_n^m(\theta, \phi)^* \,\mathrm{d}\Omega. \tag{2.92}$$

There exist two domains, the space domain, where the originating spherical function distribution $G(\mathbf{x}_\mathrm{s}, \omega)$ lives, and, the spherical wave spectrum domain (after Williams (1999)), which is the habitat of the spherical spatial Fourier coefficients $\mathring{G}_{nm}(r, \omega)$. We introduce the symbol $\mathring{\mathcal{F}}_\mathbf{x}$ for the spatial Fourier transform in a spherical coordinate system, equivalent to the symbol introduced in Eq. (2.32). Now a shorthand notation for Eq. (2.92) is given by

$$\mathring{G}_{nm}(r_0, \omega) = \mathring{\mathcal{F}}_{\mathbf{x},nm}\big\{G(\mathbf{x}_{\mathrm{s},0}, \omega)\big\}. \tag{2.93}$$

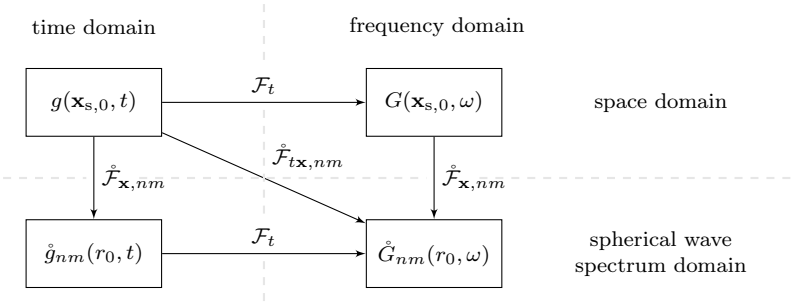


Figure 2.8 Transform relations and respective domains for the temporal and the spatial Fourier transforms in spherical coordinates. The diagram shows the forward transforms. In the corresponding scheme for the inverse transforms the arrows are reversed.

The temporal Fourier transform $\mathcal{F}_t$ is equivalent to the definition given earlier in Eq. (2.28). The same holds true for its inverse transform. Combining the spatial Fourier transform in spherical coordinates $\mathring{\mathcal{F}}_{\mathbf{x},nm}$ and the temporal Fourier transform $\mathcal{F}_t$ to a transform operation $\mathring{\mathcal{F}}_{t\mathbf{x},nm}$ yields

$$\mathring{G}_{nm}(r_0, \omega) = \iint_{S_0} \int_{-\infty}^{\infty} g(\mathbf{x}_{\mathrm{s},0}, t)\, \mathrm{e}^{-\mathrm{i}\omega\, t}\, Y_n^m(\theta, \phi)^* \,\mathrm{d}t \,\mathrm{d}\Omega. \tag{2.94}$$

The forward transform relations for spherical coordinate systems are depicted in Figure 2.8. Analogous to Eq. (2.93) the inverse spatial Fourier transform in spherical coordinates is denoted as

$$G(\mathbf{x}_{\mathrm{s},0},\omega) = \mathring{\mathcal{F}}^{-1}_{\mathbf{x},nm}\big\{\mathring{G}_{nm}(r_0,\omega)\big\}. \tag{2.95}$$

According to Eq. (2.88) the inverse transform is given by

$$G(\mathbf{x}'_{\mathrm{s},0},\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \mathring{G}_{nm}(r_0,\omega)\, Y_n^m(\theta',\phi'). \tag{2.96}$$

Note that the angles θ', ϕ', and $\mathbf{x}'_{\mathrm{s}}$ can be either identical to the previously used angles i.e. $\theta'{=}\theta$, $\phi'{=}\phi$, and $\mathbf{x}'_{\mathrm{s}}{=}\mathbf{x}_{\mathrm{s}}$, or differ from the latter, i.e. $\theta' \neq \theta$, $\phi' \neq \phi$ and $\mathbf{x}'_{\mathrm{s}} \neq \mathbf{x}_{\mathrm{s}}$. All operations related to the spatial Fourier transform discussed so far are restricted to the spherical surface S_0 with a dedicated radius r_0. This is quite obvious, since up to now the transform operation admits the expansion of arbitrary functions on the sphere without restrictions. No information on the nature of the function or field outside or inside the sphere S_0 can be assumed unless certain restrictions are introduced. In the present context of sound field analysis these restrictions are precisely determined by the Helmholtz equation given in Eq. (2.62). Assuming that all waves in a sound field follow the Helmholtz equation, required knowledge on the nature of the field outside and inside the sphere S_0 is gained. The spherical harmonics exactly solve the angular portion of the Helmholtz equation, which is discussed above. Adding the missing radial portion of the Helmholtz equation, the dedicated sphere S_0 can be left and the description expands into the full three-dimensional space. Depending on the given constellation we distinguish between interior and exterior problems.

2.13 Interior and Exterior Problems

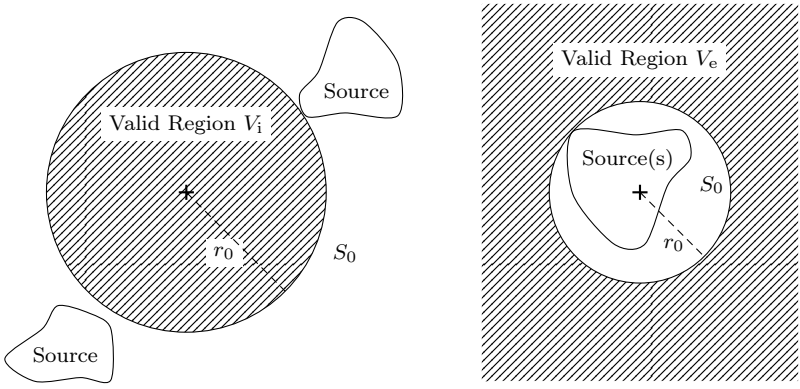
Interior and exterior problems differ from each other depending on the location of the sound sources and the resulting valid regions for sound field calculations. If all sources are surrounding the sphere S_0 the problem is called interior. The interior volume V_{i} is the valid region in that case. If the sphere S_0 encloses the sources, the problem is called exterior. The exterior volume V_{e}, extending from S_0 to infinity, is the valid region here. Both are illustrated in Figure 2.9. A third class are mixed problems, which are combinations of interior and exterior problems. Involving the radial portion of the Helmholtz equation gives access to the spatial sound field outside of S_0, as mentioned above. A new class of expansion coefficients arises that is detached from the radius r_0 and involves sufficient information to reconstruct the full three-dimensional sound

field in its corresponding region of validity V_i or V_e, respectively. In contrast to the last section where arbitrary functions on the sphere were admitted, the next expressions are firmly connected to the Helmholtz equation and wave fields. To point out this difference we directly insert a pressure distribution $P(\mathbf{x}_{\mathrm{s},0}, \omega)$ with $\mathbf{x}_{\mathrm{s},0} \in S_0$ instead of inserting any arbitrary function $G(\mathbf{x}_{\mathrm{s},0}, \omega)$. Dealing with an interior problem, all waves run from the outside region through the sphere S_0 into the volume V_i. Incoming waves are described by spherical Bessel functions of the first kind j_n. Thus, the interior expansion coefficients $\dot{P}^\mathrm{i}_{nm}(\omega)$ are calculated using (Williams, 1999, p 218)

$$\dot{P}^\mathrm{i}_{nm}(\omega) = \frac{1}{j_n\left(\frac{\omega}{c} r_0\right)} \iint_{S_0} P(\mathbf{x}_{\mathrm{s},0}, \omega)\, Y_n^m(\theta, \phi)^* \,\mathrm{d}\Omega. \tag{2.97}$$

According to Eq. (2.93) the latter can also be written as

$$\dot{P}^\mathrm{i}_{nm}(\omega) = \frac{1}{j_n\left(\frac{\omega}{c} r_0\right)} \mathring{\mathcal{F}}_{\mathbf{x},nm}\left\{P(\mathbf{x}_{\mathrm{s},0}, \omega)\right\}. \tag{2.98}$$



(a) Interior Problem (b) Exterior Problem

Figure 2.9 Two-dimensional illustration of an interior problem (a) and an exterior problem (b) (Williams, 1999). The sphere S_0 with radius r_0 is shown with surrounding sources and the maximum extension touching the sources $r_0{=}r_\mathrm{max}$ in (a), and with enclosed sources and the minimum extension touching the sources $r_0{=}r_\mathrm{min}$ in (b). The valid spatial regions for sound field calculations are hatched.

If S_0 encloses all sources, all waves emerge from the inside of S_0, pass through S_0 and run into the exterior region V_e. This is defined as exterior problem. The outgoing

waves are described by the spherical Hankel functions of the second kind $h_n^{(2)}$. Hence, the exterior expansion coefficients $\dot{P}^{\mathrm{e}}_{nm}(\omega)$ are calculated using (Williams, 1999, p 207)

$$\dot{P}^{\mathrm{e}}_{nm}(\omega) = \frac{1}{h_n\left(\frac{\omega}{c}\, r_0\right)} \iint_{S_0} P(\mathbf{x}_{\mathrm{s},0}, \omega)\, Y_n^m(\theta, \phi)^*\, \mathrm{d}\Omega. \tag{2.99}$$

Analogous to Eq. (2.98) this can also be written as

$$\dot{P}^{\mathrm{e}}_{nm}(\omega) = \frac{1}{h_n\left(\frac{\omega}{c}\, r_0\right)} \mathring{\mathcal{F}}_{\mathbf{x},nm}\big\{P(\mathbf{x}_{\mathrm{s},0}, \omega)\big\}. \tag{2.100}$$

Note that both, $\dot{P}^{\mathrm{i}}_{nm}(\omega)$ and $\dot{P}^{\mathrm{e}}_{nm}(\omega)$, are not restricted to r_0 anymore. Due to the conditions imposed by the Helmholtz equation, both contain sufficient information on the nature of the underlying field to describe the complete sound field in the respective valid regions V_{i} and V_{e}, with the exception of some specific problems that arise in Eq. (2.97) when the spherical Bessel function in the denominator goes to zero. We come back to this problem when discussing different sphere configurations.

If the expansion coefficients $\dot{P}^{\mathrm{i}}_{nm}(\omega)$ or $\dot{P}^{\mathrm{e}}_{nm}(\omega)$ are known, inverse operations can be applied for calculating the sound pressure distribution $P(\mathbf{x}_{\mathrm{s}}, \omega)$ on a sphere S with $\mathbf{x}_{\mathrm{s}} \in S \in \{V_{\mathrm{i}}, V_{\mathrm{e}}\}$. For the interior problem the inverse operation yields (Williams, 1999, p 218)

$$P(\mathbf{x}_{\mathrm{s}}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \dot{P}^{\mathrm{i}}_{nm}(\omega)\, j_n\left(\tfrac{\omega}{c}\, r\right) Y_n^m(\theta, \phi). \tag{2.101}$$

Analogous, the inverse operation for the exterior problem is given by (Williams, 1999, p 206)

$$P(\mathbf{x}_{\mathrm{s}}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \dot{P}^{\mathrm{e}}_{nm}(\omega)\, h_n\left(\tfrac{\omega}{c}\, r\right) Y_n^m(\theta, \phi). \tag{2.102}$$

The important detail is the radius r that can either be identical to r_0 or can generally be defined as $r \leq r_0$ for interior problems or $r \geq r_0$ for exterior problems. Knowing the expansion coefficients, the sound field can be calculated on an arbitrary sphere S with radius r that lies in the corresponding valid region V_{i} or V_{e}.

2.14 Sound Field Extrapolation in Spherical Coordinates

The basic idea of sound field extrapolation is introduced for Cartesian coordinates in Section 2.8. For spherical coordinates the concept is the same but the procedure

is different. Combining forward and inverse transforms appropriately, the pressure distribution on a sphere S_0 with radius r_0 can be related to the pressure distribution on a sphere S' with radius r' in the space domain. For an interior problem the following equation arises (Williams, 1999, p 218):

$$P(\mathbf{x}_\mathrm{s}', \omega) = \sum_{n=0}^{\infty} \frac{j_n\left(\frac{\omega}{c}\, r'\right)}{j_n\left(\frac{\omega}{c}\, r_0\right)} \sum_{m=-n}^{n} Y_n^m(\theta', \phi') \iint_{S_0} P(\mathbf{x}_{\mathrm{s},0}, \omega)\, Y_n^m(\theta, \phi)^* \mathrm{d}\Omega, \quad (2.103)$$

where the vector $\mathbf{x}_\mathrm{s}' \in S'$ involves θ', ϕ', and r', and $\mathbf{x}_{\mathrm{s},0} \in S_0$ involves θ, ϕ, and r_0. The spatial Fourier transform can be written using the previously defined transform operator $\mathring{\mathcal{F}}_{\mathbf{x},nm}$ yielding

$$P(\mathbf{x}_\mathrm{s}', \omega) = \sum_{n=0}^{\infty} \frac{j_n\left(\frac{\omega}{c}\, r'\right)}{j_n\left(\frac{\omega}{c}\, r_0\right)} \sum_{m=-n}^{n} Y_n^m(\theta', \phi')\; \mathring{\mathcal{F}}_{\mathbf{x},nm}\big\{P(\mathbf{x}_{\mathrm{s},0}, \omega)\big\}. \quad (2.104)$$

For an exterior problem with $r' \geq r_0 \geq r_{0,\min}$ the following equation arises (Williams, 1999, p 207):

$$P(\mathbf{x}_\mathrm{s}', \omega) = \sum_{n=0}^{\infty} \frac{h_n\left(\frac{\omega}{c}\, r'\right)}{h_n\left(\frac{\omega}{c}\, r_0\right)} \sum_{m=-n}^{n} Y_n^m(\theta', \phi') \iint_{S_0} P(\mathbf{x}_{\mathrm{s},0}, \omega)\, Y_n^m(\theta, \phi)^* \mathrm{d}\Omega, \quad (2.105)$$

which can be written as

$$P(\mathbf{x}_\mathrm{s}', \omega) = \sum_{n=0}^{\infty} \frac{h_n\left(\frac{\omega}{c}\, r'\right)}{h_n\left(\frac{\omega}{c}\, r_0\right)} \sum_{m=-n}^{n} Y_n^m(\theta', \phi')\; \mathring{\mathcal{F}}_{\mathbf{x},nm}\big\{P(\mathbf{x}_{\mathrm{s},0}, \omega)\big\}. \quad (2.106)$$

As an alternative, the extrapolation can be performed in the spherical wave spectrum domain by directly relating spatial Fourier coefficients. If we insert a pressure distribution $P(\mathbf{x}_\mathrm{s}, \omega)$ into Eq. (2.93) we obtain spatial Fourier coefficients $\mathring{P}_{nm}(r, \omega)$ in the spherical wave spectrum domain:

$$\mathring{P}_{nm}(r, \omega) = \mathring{\mathcal{F}}_{\mathbf{x},nm}\big\{P(\mathbf{x}_\mathrm{s}, \omega)\big\}, \quad (2.107)$$

For interior problems with $r' \leq r_0$ the relation between coefficients corresponding to different radii in the spherical wave spectrum domain yields (Williams, 1999, p 218)

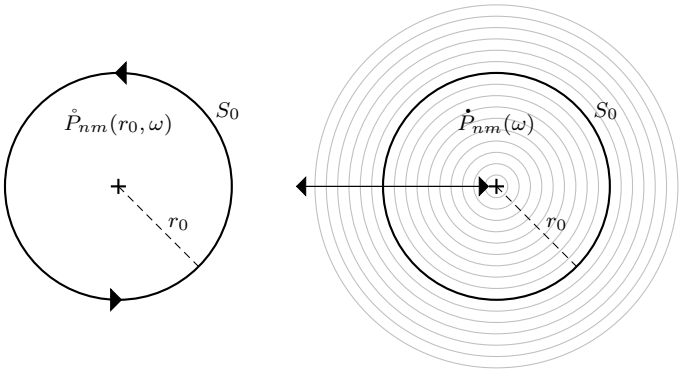
$$\mathring{P}_{nm}(r', \omega) = \frac{j_n\left(\frac{\omega}{c}\, r'\right)}{j_n\left(\frac{\omega}{c}\, r_0\right)}\; \mathring{P}_{nm}(r_0, \omega). \quad (2.108)$$

For exterior problems with $r' \geq r_0$ we have (Williams, 1999, p 208)

$$\mathring{P}_{nm}(r', \omega) = \frac{h_n\left(\frac{\omega}{c}\, r'\right)}{h_n\left(\frac{\omega}{c}\, r_0\right)}\, \mathring{P}_{nm}(r_0, \omega). \tag{2.109}$$

2.15 Inter- and Extrapolation in the Spherical Wave Spectrum Domain

We need to elaborate the two types of Fourier coefficients introduced in the last sections. Eq. (2.93) can be applied to a pressure distribution $P(\mathbf{x}_{\mathrm{s},0}, \omega)$ on S_0 in order to obtain spatial Fourier coefficients $\mathring{P}_{nm}(r_0, \omega)$ in the spherical wave spectrum domain. As already discussed in Section 2.12 these coefficients do only contain information on the sound pressure distribution on the sphere S_0 and do not suffice to calculate anything outside or inside S_0, since we are not firmly tied to the full Helmholtz equation so far. Thus, at first sight these coefficients appear to be intermediate objects. However, looking at the inverse spatial Fourier transform in Eq. (2.96) it becomes apparent that the pressure can be calculated for an arbitrary point in S_0.



(a) Interpolation (b) Extrapolation

Figure 2.10 Two-dimensional illustration of (a) interpolation on the sphere S_0 using the Fourier coefficients $\mathring{P}_{nm}(r_0, \omega)$, and (b) extrapolation into the space outside of S_0 using the Fourier expansion coefficients $\dot{P}_{nm}(\omega)$ along concentric spheres. The latter does not specify whether the underlying problem is an interior or an exterior one. The valid regions have to be considered, refer to Figure 2.9, and the appropriate coefficients $\dot{P}_{nm}(\omega)$=$\dot{P}^{\mathrm{i}}_{nm}(\omega)$ for interior or $\dot{P}_{nm}(\omega)$=$\dot{P}^{\mathrm{e}}_{nm}(\omega)$ for exterior problems, respectively, need to be chosen.

If we now consider not knowing the continuous pressure distribution on S_0 but knowing the pressure at discrete sampling positions located on S_0 only, the spatial Fourier transform reveals powerful capabilities for interpolation on the sphere. This is shown in Figure 2.10(a).

In a next step, the angular portion of the Helmholtz equation is applied, as discussed in Section 2.13. Depending on the type of the problem, specific Fourier expansion coefficients are calculated which are firmly tied to the Helmholtz equation. If we have an interior problem the inner expansion coefficients $\dot{P}^{\mathrm{i}}_{nm}(\omega)$ are calculated using Eq. (2.97). If the problem is an exterior one, the exterior expansion coefficients $\dot{P}^{\mathrm{i}}_{nm}(\omega)$ are calculated using Eq. (2.99). The coefficients $\dot{P}_{nm}(\omega)$ can be used to leave the sphere S_0 and calculate the sound field on an arbitrary concentric sphere inside or outside S_0 within the valid region defined by the given problem, see Figure 2.9.

Both techniques, i.e. interpolation and extrapolation, are illustrated in Figure 2.10. An example for practical application of interpolation is described in the context of head-related transfer functions, refer to Section 3.12.4.

2.16 Rotations in the Spherical Wave Spectrum Domain

The function defined by the Fourier coefficients $\mathring{G}_{nm}(r_0, \omega)$ can be rotated on the sphere. The rotation in the spherical wave spectrum domain is quite useful, especially for rotating discretely sampled pressure distributions, since this operation provides inherent capabilities for interpolation on the sphere.

2.16.1 Rotation Group SO(3) and Wigner-D Functions

An element ν of the rotation group SO(3) (Arfken and Weber, 2005, p 250) can be expressed as a product of single rotation operators. A common convention is the z-y-z rotation, where first a rotation around the z-axis is performed, followed by a rotation around the y-axis and by another rotation around the z-axis. All rotations are counterclockwise. They are expressed using the Euler angles α, β and γ and described by the well-known 3×3 rotation matrices (Arfken and Weber, 2005, p 202)

$$\mathcal{R}_{\mathrm{z}}(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.110}$$

$$\mathcal{R}_{\mathrm{y}}(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}, \tag{2.111}$$

and

$$\mathcal{R}_{\mathrm{z}}(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.112}$$

The full rotation operator $\mathcal{R}$ on SO(3) is

$$\mathcal{R}_{\mathrm{zyz}}(\alpha,\beta,\gamma) = \mathcal{R}_{\mathrm{z}}(\alpha)\,\mathcal{R}_{\mathrm{y}}(\beta)\,\mathcal{R}_{\mathrm{z}}(\gamma). \tag{2.113}$$

Wigner-D functions $D^n_{mm'}(\alpha,\beta,\gamma)$ can be used to perform rotations in the spherical wave spectrum domain. They were introduced by Wigner (1931) and are commonly used in quantum mechanics for the description of particles with spin (Edmonds, 1957). In analogy to the rotation operator, Wigner-D functions contain single rotations with the Euler angles (Morrison and Parker, 1987):

$$D^n_{mm'}(\alpha,\beta,\gamma) = \mathrm{e}^{-\mathrm{i}\,m\,\alpha}\; d^n_{mm'}(\beta)\; \mathrm{e}^{-\mathrm{i}\,m'\,\gamma}, \tag{2.114}$$

where $d^n_{mm'}(\beta)$ denotes the Wigner-d function that can be explicitly expressed in terms of the Jacobi polynomials (Morrison and Parker, 1987):

$$d^n_{mm'}(\beta) = \sqrt{\frac{(n-m')!(n+m')!}{(n+m)!(n-m)!}}\,\big(\cos\tfrac{\beta}{2}\big)^{m'+m}\big(\sin\tfrac{\beta}{2}\big)^{m'-m} P^{(m'-m,m'+m)}_{n-m'}(\cos\beta). \tag{2.115}$$

The Jacobi polynomials in Eq. (2.115) can be written as (Morrison and Parker, 1987)

$$P^{(k_1,k_2)}_k(\cos\beta) = (k+k_1)!(k+k_2)!\times \sum_{\mathrm{s}} \frac{1}{s!(k+k_1-s)!(k_2+s)!(k-s)!}\Big(-\sin^2\frac{\beta}{2}\Big)^{k-s}\Big(\cos^2\frac{\beta}{2}\Big)^{s}, \tag{2.116}$$

with integers k, k_1 and k_2; the summation involves all integers for s that have positive arguments of the factorials. Wigner-D functions form a complete orthogonal (but not orthonormal) set of functions with respect to integration over the rotation group SO(3). The orthogonality relation yields (Edmonds, 1957, p 62)

$$\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} D^{n_2}_{m_2m_2'}(\alpha,\beta,\gamma)^*\; D^n_{mm'}(\alpha,\beta,\gamma)\;\mathrm{d}\alpha\;\sin\beta\,\mathrm{d}\beta\,\mathrm{d}\gamma = \frac{8\pi^2}{2n+1}\delta_{m_2m_1}\,\delta_{m_2'm_1'}\,\delta_{n_2n_1}. \tag{2.117}$$

Due to the orthogonality, an arbitrary function $F(\alpha, \beta, \gamma) \in L^2(SO(3))$ can be decomposed using Wigner-D functions. Spherical harmonics are related to the Wigner-D functions for $m' = 0$ according to (Pendleton, 2003)

$$D^n_{m0}(\alpha, \beta, \gamma) = \sqrt{\frac{4\pi}{2n+1}} Y^m_n(\beta, \alpha)^*. \tag{2.118}$$

The Wigner-D functions form matrix entries for the spherical rotation operator $\mathring{\mathcal{R}}(\alpha, \beta, \gamma)$ in the spherical wave spectrum domain and can be used for rotation of the spherical harmonics yielding (Morrison and Parker, 1987)

$$Y^{m'}_n(\theta', \phi') = \sum_{m=-n}^{n} Y^m_n(\theta, \phi)\, D^n_{mm'}(\alpha, \beta, \gamma). \tag{2.119}$$

For complex conjugate spherical harmonics the equation reads (Varshalovich et al., 1988, p 72)

$$Y^{m'}_n(\theta', \phi')^* = \sum_{m=-n}^{n} Y^m_n(\theta, \phi)^*\, D^n_{mm'}(\alpha, \beta, \gamma)^*. \tag{2.120}$$

θ' and ϕ' describe the new angles after rotation. As an alternative, the rotation in Eq. (2.119) can be expressed in terms of a rotation operator $\mathring{\mathcal{R}}_{\mathrm{zyz}}(\alpha, \beta, \gamma)$ in the spherical wave spectrum domain:

$$\mathring{\mathcal{R}}_{\mathrm{zyz}}(\alpha, \beta, \gamma)\, Y^{m'}_n(\theta, \phi) = \sum_{m=-n}^{n} Y^m_n(\theta, \phi)\, D^n_{mm'}(\alpha, \beta, \gamma). \tag{2.121}$$

This is equivalent to the operator $\mathcal{R}_{\mathrm{zyz}}(\alpha, \beta, \gamma)$ in the space domain introduced in Eq. (2.113). We formulate a function $G(\mathbf{x}_{\mathrm{s},0}, \omega)$ on the sphere in the space domain. This function can be rotated by applying the rotation operator $\mathcal{R}_{\mathrm{zyz}}(\alpha, \beta, \gamma)$ as defined in Eq. (2.113). The rotated function is denoted as $G^{\mathcal{R}}(\mathbf{x}_{\mathrm{s},0}, \omega)$:

$$G^{\mathcal{R}}(\mathbf{x}_{\mathrm{s},0}, \omega) = \mathcal{R}_{\mathrm{zyz}}(\alpha, \beta, \gamma)\, G(\mathbf{x}_{\mathrm{s},0}, \omega)\,. \tag{2.122}$$

In Eq. (2.96) the expansion of $G(\theta, \phi)$ in spherical harmonics is given. Applying Eq. (2.121) to this expansion, where m is changed for m', the rotation can be expressed as

$$G^{\mathcal{R}}(\mathbf{x}_{\mathrm{s},0}, \omega) = \mathcal{R}_{\mathrm{zyz}}(\alpha, \beta, \gamma)\, G(\mathbf{x}_{\mathrm{s},0}, \omega) \tag{2.123}$$

$$= \sum_{n=0}^{\infty} \sum_{m'=-n}^{n} \mathring{G}_{nm'}(r_0, \omega)\, \mathring{\mathcal{R}}_{\mathrm{zyz}}(\alpha, \beta, \gamma)\, Y_n^{m'}(\theta, \phi) \tag{2.124}$$

$$= \sum_{n=0}^{\infty} \sum_{m'=-n}^{n} \mathring{G}_{nm'}(r_0, \omega) \sum_{m=-n}^{n} Y_n^m(\theta, \phi)\, D_{mm'}^n(\alpha, \beta, \gamma) \tag{2.125}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[\sum_{m'=-n}^{n} \mathring{G}_{nm'}(r_0, \omega)\, D_{mm'}^n(\alpha, \beta, \gamma) \right] Y_n^m(\theta, \phi) \tag{2.126}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}_{nm}^{\mathcal{R}}(r_0, \omega)\, Y_n^m(\theta, \phi). \tag{2.127}$$

This is similar to the expression in (Rafaely and Kleider, 2008). It leads to the conclusion that the spatial Fourier coefficients $\mathring{G}_{nm}$ corresponding to $G(\theta, \phi)$ can be rotated in the spherical wave spectrum domain using

$$\mathring{G}_{nm}^{\mathcal{R}} = \sum_{m'=-n}^{n} \mathring{G}_{nm'}(r_0, \omega)\, D_{mm'}^n(\alpha, \beta, \gamma). \tag{2.128}$$

Other aspects of rotations and Wigner-D functions are discussed e.g. in (Edmonds, 1957, pp 53–65), (Varshalovich et al., 1988, pp 72–117), (Morrison and Parker, 1987) or (Pendleton, 2003). Wigner-D functions for array beam steering are discussed in (Rafaely and Kleider, 2008), the use of the same for rotating directivity patterns in room acoustic simulations are treated in (Pelzer et al., 2012). Wigner-D functions for improving matrix conditioning in the context of direction of arrival (DOA) estimation with microphone arrays are discussed in (Sun et al., 2011).

2.16.2 Euler Rotation

For several applications only rotations around the z-axis are required. A reduced rotation operator, called Euler rotation, is introduced in order to eliminate the overhead that comes with the full Wigner-D functions in this special case. For $\{\beta, \gamma\}$=0 the Wigner-D function $D_{mm'}^n(\alpha, \beta, \gamma)$ reduces to a function $D_m(\alpha)$, which according to Eq. (2.114) is defined as

$$D_m(\alpha) = \mathrm{e}^{-\mathrm{i}m\alpha}, \tag{2.129}$$

where α denotes the rotation around the z-axis. The corresponding Euler rotation matrix in the space domain is given by Eq. (2.110). Analogous to Eq. (2.119), the rotation of spherical harmonics yields

$$Y_n^m(\theta, \phi') = Y_n^m(\theta, \phi)\, D_m(\alpha). \tag{2.130}$$

According to Eq. (2.121) this can be expressed in terms of a reduced rotation operator $\mathring{\mathcal{R}}_z(\alpha)$ as

$$\mathring{\mathcal{R}}_z(\alpha)\, Y_n^m(\theta, \phi) = Y_n^m(\theta, \phi)\, D_m(\alpha). \tag{2.131}$$

The equivalent approach to Eq. (2.123) to Eq. (2.127), written as

$$\begin{aligned}
G^{\mathcal{R}}(\mathbf{x}_{s,0}, \omega) &= \mathcal{R}_z(\alpha)\, G(\mathbf{x}_{s,0}, \omega) && (2.132)\\
&= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}_{nm}(r_0, \omega)\, \mathring{\mathcal{R}}_z(\alpha)\, Y_n^m(\theta, \phi) && (2.133)\\
&= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}_{nm}(r_0, \omega)\, Y_n^m(\theta, \phi)\, D_m(\alpha) && (2.134)\\
&= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[\mathring{G}_{nm}(r_0, \omega)\, D_m(\alpha)\right] Y_n^m(\theta, \phi) && (2.135)\\
&= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}^{\mathcal{R}}_{nm}(r_0, \omega)\, Y_n^m(\theta, \phi), && (2.136)
\end{aligned}$$

indicates that the reduced rotation of spatial Fourier coefficients $\mathring{G}_{nm}$ corresponding to $G(\theta, \phi)$ can be expressed as

$$\mathring{G}^{\mathcal{R}}_{nm} = \mathring{G}_{nm}(r_0, \omega)\, D_m(\alpha). \tag{2.137}$$

The result is a simple operation with low computational demands in practical applications.

2.17 Plane Wave Expansion

The plane wave is one of the most important concepts when talking about waves. In the present work the plane wave plays a major role, since the objective is to decompose

the sound fields into plane wave components. In theory, plane waves are assumed to arise from a point source that is located at infinite distance. The curvature of the wave front vanishes before the wave reaches the observation point. The plane wave solution to the Helmholtz equation in Cartesian coordinates is introduced in Eq. (2.47) and, inverting the propagation direction for convenience, an incident plane wave reads

$$P_{\mathrm{pw}}(\mathbf{x}_{\mathrm{c}}, \omega) = \widehat{P}\, \mathrm{e}^{\mathrm{i}\, \mathbf{k}_{\mathrm{c,pw}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c}}}. \tag{2.138}$$

If the position vector is expressed in spherical coordinates according to Eq. (2.56), the plane wave can be simply written as

$$P_{\mathrm{pw}}(\mathbf{x}_{\mathrm{s}}, \omega) = \widehat{P}\, \mathrm{e}^{\mathrm{i}\, \mathbf{k}_{\mathrm{s,pw}}^{\mathrm{T}} \mathbf{x}_{\mathrm{s}}}, \tag{2.139}$$

where the wave vector $\mathbf{k}_{\mathrm{s,pw}}^{\mathrm{T}}$ points into the direction of wave propagation. It is given according to a k_{pw}-weighted version of the unit vector $\vec{\mathbf{e}}_{\mathrm{r}}$ in Eq. (2.57) by $\mathbf{k}_{\mathrm{s,pw}}^{\mathrm{T}} = k_{\mathrm{pw}} \left[\sin\theta_{\mathrm{pw}} \cos\phi_{\mathrm{pw}} \;\; \sin\theta_{\mathrm{pw}} \sin\phi_{\mathrm{pw}} \;\; \cos\theta_{\mathrm{pw}}\right]^{\mathrm{T}}$, with θ_{pw} and ϕ_{pw} denoting the wave arrival direction observed from the origin and k_{pw} the respective wave number similar to (Ahrens, 2010, p 4). The pressure distribution on S_0 that is generated by the incident plane wave can be expressed in terms of spherical harmonics according to Eq. (2.101), which is shown in (Williams, 1999, pp 225–227). The expansion of an incident plane wave arriving from $(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})$ yields (Williams, 1999, p 227), (Gumerov and Duraiswami, 2004, p 74)

$$P_{\mathrm{pw}}(\mathbf{x}_{\mathrm{s}}, \omega) = \widehat{P}\, 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathrm{i}^n j_n\left(\tfrac{\omega}{c}\, r\right) Y_n^m(\theta, \phi)\, Y_n^m(\theta_{\mathrm{w}}, \phi_{\mathrm{w}})^*. \tag{2.140}$$

2.18 Spherical Waves in Spherical Coordinates

The plane wave assumption is not valid for the observation of point sources located in the near field, as the curvature of the wave front is not negligible anymore. The sound pressure P_{sw} at the location $\mathbf{x}_{\mathrm{s}}$ produced by a monopole source located at $\mathbf{x}_{\mathrm{s}}' = r'\, \vec{\mathbf{e}}_{r'} + \theta'\, \vec{\mathbf{e}}_{\theta'} + \phi\, \vec{\mathbf{e}}_{\phi'}$ is described by (Williams, 1999, p 259)

$$P_{\mathrm{sw}}(\mathbf{x}_{\mathrm{s}}, \mathbf{x}_{\mathrm{s}}', \omega) = \widehat{P}\, \frac{e^{\mathrm{i}\, \frac{\omega}{c}\, |\mathbf{x}_{\mathrm{s}}' - \mathbf{x}_{\mathrm{s}}|}}{|\mathbf{x}_{\mathrm{s}}' - \mathbf{x}_{\mathrm{s}}|}, \tag{2.141}$$

which, besides a normalization factor of $1/4\pi$, is known as free-field Green's function. For $r_0 < r'$, where the monopole source in $\mathbf{x}_{\mathrm{s}}'$ is further away from the origin than the observation point $\mathbf{x}_{\mathrm{s}}$, the expansion of Eq. (2.141) in spherical harmonics yields (Williams, 1999, p 259)

$$P_{\mathrm{sw}}(\mathbf{x}_{\mathrm{s}}, \mathbf{x}'_{\mathrm{s}}, \omega) = \widehat{P}\, 4\pi\, \mathrm{i}\, \frac{\omega}{c} \sum_{n=0}^{\infty} j_n\big(\tfrac{\omega}{c}\, r_0\big)\, h_n\big(\tfrac{\omega}{c}\, r'\big) \times \sum_{m=-n}^{n} Y_n^m(\theta', \phi')\, Y_n^m(\theta, \phi)^*. \tag{2.142}$$

For this thesis, dedicated near field observations are neglected. The sound sources are assumed to be located in the far field and the focus of interest is on the decomposition of plane waves.

2.19 Plane Wave Decomposition (PWD)

Analogous to the plane wave expansion discussed in Section 2.17, a given sound field can be decomposed into its plane wave components. According to the principle of superposition stated in Section 2.1, the decomposition is valid for both, single plane waves and arbitrarily complex sound fields. The latter can be described by superimposed plane waves. We assume a continuous pressure distribution $P(\theta, \phi, r_0, \omega)$ on S_0 with radius r_0. Its corresponding spatial Fourier coefficients are denoted by $\mathring{P}_{nm}(r_0, \omega)$. Both are related through the spatial Fourier transform, cf. Section 2.92. The plane wave decomposition (PWD) (Rafaely, 2004), (Duraiswami et al., 2005a) returns the plane wave components D for a specific spatial decomposition direction $(\theta_{\mathrm{d}}, \phi_{\mathrm{d}})$ yielding

$$D(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n\, j_n(\tfrac{\omega}{c}\, r_0)}\, \mathring{P}_{nm}(r_0, \omega)\, Y_n^m(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}). \tag{2.143}$$

PWD can be understood as a spatial Dirac pulse pointing into direction $(\theta_{\mathrm{d}}, \phi_{\mathrm{d}})$. This enables ideal spatial sampling in terms of plane waves. The Dirac pulse is ideal, as long as the sum $n = [0, \infty[$ is infinite. The consequences of truncating the sum to a limited order N, $n = [0, N]$ are discussed in Section 3.3.

At this point we still rely on a continuous and error-free description of the sound field on S_0, which would indeed enable perfect PWD. As soon as technical systems, i.e. microphone arrays, are used to capture the information on S_0, the PWD looses several of its ideal properties. Section 3 is dedicated to analyzing and discussing the respective constraints in technical systems.

PWD is a special case of modal beamforming using what is referred to as regular beam pattern (Li and Duraiswami, 2007). This term becomes more explicit in Section 3.3, where the spatial Dirac pulse degenerates due to the truncation of the modal order. In order to perform beamforming with arbitrarily shaped beams, specific beamforming

coefficients $\mathring{B}_{nm}$ can be introduced to Eq. (2.143), which establish the target beam shape:

$$D^B(\theta_\mathrm{d}, \phi_\mathrm{d}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{\mathring{B}_{nm}(\omega)}{\mathrm{i}^n \, j_n(\frac{\omega}{c} r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta_\mathrm{d}, \phi_\mathrm{d}). \tag{2.144}$$

The PWD provides constant coefficients $\mathring{B}_{nm} = 1$, cf. Eq. 2.143. Different approaches and applications of modal beamforming are not further discussed. The reader is referred e.g. to (Teutsch, 2007) for a detailed discussion of modal beamforming.

2.20 Binaural Reproduction of Physical Sound Field Descriptions

For binaural playback of sound fields that are described in the modal domain, the modal sound field description needs to be merged with HRTFs. For that purpose, the sound field in $\mathring{P}_{nm}(r_0, \omega)$ is decomposed into separate directional contributions $(\theta, \phi) \in S$ using PWD according to Eq. (2.143). Then the separate directional output signals are weighted with corresponding complex-valued HRTFs[1] with angular directions $(\theta, \phi) \in S$. Finally, all directional HRTF-weighted signals are integrated over the complete sphere S in order to deliver a binaural output signal $Y^{\mathrm{l,r}}(\omega)$. This operation yields:

$$Y^{\mathrm{l,r}}(\omega) = \frac{1}{4\pi} \iint_S H^{\mathrm{l,r}}(\theta, \phi, \omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c} r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta, \phi) \, \mathrm{d}\Omega. \tag{2.145}$$

In order to justify this approach, we define a basic analytic wave field in the spherical wave spectrum domain that consists of a single plane wave with unit amplitude arriving from direction $(\theta_\mathrm{w}, \phi_\mathrm{w})$ (Williams, 1999, p 259):

$$\mathring{P}_{nm\,\mathrm{pw}(\theta_\mathrm{w}, \phi_\mathrm{w})}(r_0, \omega) = 4\,\pi\,\mathrm{i}^n \, j_n(\tfrac{\omega}{c} r_0) \, Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^*. \tag{2.146}$$

In a next step, analogous to Eq. (2.145), we perform a single PWD and weight the output signal with the corresponding HRTF yielding

$$Y^{\mathrm{l,r}\,\prime}(\theta, \phi, \omega) = H^{\mathrm{l,r}}(\theta, \phi, \omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c} r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta, \phi). \tag{2.147}$$

[1] Corresponding to the convolution in the time-space domain.

When replacing the general sound field description $\mathring{P}_{nm}(r_0, \omega)$ in Eq. (2.147) by the analytic plane wave description $\mathring{P}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0, \omega)$ from Eq. (2.146), the relation

$$Y^{\mathrm{l,r}\,\prime}(\theta, \phi, \omega) = H^{\mathrm{l,r}}(\theta, \phi, \omega)\, 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\theta, \phi)\, Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^* \tag{2.148}$$

emerges, which means

$$Y^{\mathrm{l,r}\,\prime}(\theta, \phi, \omega) = \begin{cases} H^{\mathrm{l,r}}(\theta, \phi, \omega) & \text{if } (\theta_\mathrm{w}, \phi_\mathrm{w}) = (\theta, \phi), \text{ and} \\ 0 & \text{if } (\theta_\mathrm{w}, \phi_\mathrm{w}) \neq (\theta, \phi). \end{cases} \tag{2.149}$$

The spatial Dirac pulse in the expression noted above sifts out plane wave incidence that matches the original HRTF direction.

Inserting the analytic plane wave description $\mathring{P}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0, \omega)$ from Eq. (2.146) into Eq. (2.145) yields

$$Y^{\mathrm{l,r}}_{\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(\omega) = \iint_S H^{\mathrm{l,r}}(\theta, \phi, \omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\theta, \phi)\, Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^*\, \mathrm{d}\Omega. \tag{2.150}$$

For the specific case of a single plane wave incidence from the direction $(\theta_\mathrm{w}, \phi_\mathrm{w})$ this becomes

$$Y^{\mathrm{l,r}}_{\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(\omega) = H^{\mathrm{l,r}}(\theta_\mathrm{w}, \phi_\mathrm{w}, \omega). \tag{2.151}$$

Assuming ideal conditions, the operation actually turns out to be transparent for plane wave incidence. Due to the principle of superposition, the transparency property can be generalized to complex sound fields.

The real advantage of this approach is the ability to deduce a (theoretically) ideal binaural signal from the knowledge of the sound field properties on a static sphere S_0, while free head rotation and even translation can be performed in a separate stage.

Rotation can be applied in different ways. One option is to rotate the entire sound sound field $P_{nm}(r_0, \omega)$ in Eq. (2.145) around the angles α, β and γ using Wigner-D functions (cf. Section 2.16.1), yielding rotated Fourier coefficients $\mathring{P}^{\mathcal{R}}_{nm}(r_0, \omega)$ with

$$\mathring{P}^{\mathcal{R}}_{nm}(r_0, \omega) = \sum_{m'=-n}^{n} \mathring{P}_{nm'}(r_0, \omega)\, D^n_{mm'}(\alpha, \beta, \gamma)\,. \tag{2.152}$$

A different option is to apply a pre-rotated HRTF set $H^{\mathrm{l,r}}(\theta', \phi', \omega)$ instead of $H^{\mathrm{l,r}}(\theta, \phi, \omega)$ in Eq. (2.145), where (θ', ϕ') describes the rotated angle pair. If not explicitly available, rotated HRTFs $H^{\mathrm{l,r}}(\theta', \phi', \omega)$ can be generated using spatial Fourier transforms and Wigner-D functions. A spherical HRTF set $H^{\mathrm{l,r}}(\theta, \phi, \omega)$ can be regarded as a frequency-dependent description of a closed complex function with respect to magnitude and phase on a virtual sphere S. Thus, the spatial Fourier transform (cf. Eq. (2.92)) is directly applicable:

$$\mathring{H}^{\mathrm{l,r}}_{nm}(\omega) = \iint_S H^{\mathrm{l,r}}(\theta, \phi, \omega)\, Y^m_n(\theta, \phi)^*\, \mathrm{d}\Omega. \tag{2.153}$$

In a next step, Wigner-D rotation (cf. Section 2.16.1) around the angles α, β and γ is applied, yielding a rotated HRTF set $\mathring{H}^{\mathrm{l,r}\,\mathcal{R}}_{nm}(r_0, \omega)$,

$$\mathring{H}^{\mathrm{l,r}\,\mathcal{R}}_{nm}(\omega) = \sum_{m'=-n}^{n} \mathring{H}^{\mathrm{l,r}}_{nm'}(\omega)\, D^n_{mm'}(\alpha, \beta, \gamma). \tag{2.154}$$

In a last step, the inverse spatial Fourier transform (cf. Eq. (2.96)) is used to achieve a rotated HRTF $H^{\mathrm{l,r}}(\theta', \phi', \omega)$ in the space-frequency domain related to the original angle pair (θ, ϕ),

$$H^{\mathrm{l,r}}(\theta', \phi', \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{H}^{\mathrm{l,r}\,\mathcal{R}}_{nm}(\omega)\, Y^m_n(\theta, \phi). \tag{2.155}$$

As an alternative to applying Wigner-D functions for rotating the HRTFs, the angle offset can be directly applied to (θ, ϕ) during the backward Fourier transform in Eq. (2.155).

This operation might appear excessively complicated at first. Nevertheless, the rotation of the HRTF set usually requires less computational power than rotating the entire sound field description. Furthermore, forward and backward spatial Fourier transforms inherently performs spherical harmonic HRTF interpolation, which is of particular interest for discretely sampled HRTF sets that are used in practice. Discrete spatial sampling of HRTFs and spherical harmonic HRTF interpolation are discussed in Section 3.12. Thus the benefit of this approach becomes more explicit later.

An additional option is translating the head within the source-free region of the sound field. This is done by successively extrapolating the Fourier coefficients $\mathring{P}_{nm'}(r_0, \omega)$ to a shifted sphere using Eq. (2.108). The basic approach for sound field extrapolation is discussed in Section 2.14.

The entire theoretical approach clearly unveils a certain grace and completeness. A transparent binaural system providing full horizontal, vertical and lateral head rotation and even theoretical head translation capabilities that are applicable in a separate post-processing stage can be designed. Nevertheless, several of the previously assumed ideal conditions cannot be established in technical systems.

3 Constraints in Technical Systems

3.1 Discretization of Time and Amplitude

Processing of the audio signals is performed with digital systems. As a consequence, the signals are discretized with respect to time and amplitude. Discretization of time and amplitude does not have any special impact on the presented approaches. The reader is referred to the extensive available literature on digital audio processing, e.g. (Kefauver and Patschke, 2007) or (Zölzer, 2008).

In order to simplify the notation, signals are denoted as ideal signals with continuous time and amplitude resolution, even though in practice signals with discrete time and amplitude resolution are addressed. The corresponding discrete signals are assumed to be band-limited to frequencies below half of the temporal sampling rate and to have sufficient amplitude resolution. Thus, there is no relevant restriction of the dynamic range in the systems and no spectro-temporal aliasing is present according to the Nyquist-Shannon sampling theorem (Nyquist, 1928), (Shannon, 1949). Since the signals are denoted as continuous signals, the continuous time-frequency Fourier transform is used in the equations, according to the definitions from Section 2.4. When processing discrete signals in practice, the continuous Fourier transforms need to be replaced with appropriate discrete transforms such as the discrete Fourier transform (DFT) or the fast Fourier transform (FFT), cf. (Jackson, 1996).

Unless otherwise specified, a temporal sampling rate of 48 kHz is used for all measurements and simulations throughout this work. All signals are band-limited to 20 kHz. Signal conversions between the analog and the digital domain are performed using analog to digital converters (ADCs) or digital to analog converters (DACs) with a word length of 24 bit. Computations are performed using double precision floating point variables. Time-frequency Fourier transforms are based on FFT algorithms.

3.2 Spatial Discretization

Up to this point we assumed to have continuous information on the sound pressure on the entire sphere S_0. Currently, no known technical system is capable of acquiring sound field information in real-time that is even close to a continuous description. Future technologies, e.g. based on floating air-filled spherical membranes scanned with

a laser[1] might possibly come much closer to this goal. In the meantime, microphone arrays with their membranes located on the sphere S_0 are used for the acquisition of the sound pressure in practice. As a consequence of using microphone arrays, the pressure distribution on S_0 is sampled at a limited amount of discrete spatial sampling nodes, which severely impacts the transmission properties of the previously defined ideal system. This raises basic questions concerning the expected consequences of discrete spatial sampling, the required number of sampling nodes, or the reasonable distribution of nodes on the sphere.

3.2.1 Microphone Arrays

Classical real-time microphone arrays carry a microphone at each spatial sampling node and provide an individual signal path including pre-amplifier and ADC. Examples of classical real-time arrays that are specifically made for modal analysis of exterior sound fields are the popular EIGENMIKE® from mh-Acoustics (Meyer and Elko, 2002) or several experimental arrays that were built in the scientific community such as the arrays presented in (Li and Ruraiswami, 2005), (Duraiswami et al., 2005b), (O'Donovan et al., 2008), or (Peters and Schmeder, 2011).

A particular type of arrays is called scanning array or virtual array. The scanning array carries only one microphone (or few microphones) mounted on a robotic arm, which can be moved to different spatial positions. The signals for each transducer position of the array are acquired sequentially. This procedure is feasible, as long as the surroundings behave sufficiently time-invariant. Scanning arrays are usually employed for acquiring (room) impulse responses. In contrast to classical microphone arrays, this kind of array is generally not suitable for capturing sound fields in real-time. Nevertheless, the advantage of scanning arrays is the large flexibility concerning the positioning and number of sampling points, which is highly convenient for research purposes. Very dense sampling grids including a nearly arbitrary amount of sampling positions on different radii can be captured for a single array position, which is hardly possible using conventional microphone arrays according to the current state of technology. Since a spherical microphone only uses a single microphone and the identical audio path for all sampling nodes, no specific array calibration (Rettberg et al., 2012) is necessary. A first spherical scanning microphone array was developed by Schlesinger et al. (2007). Another spherical scanning array is used by Rafaely et al. (2007a). For this thesis a scanning microphone array system called VariSphear was developed, which is described in Section 4.1.

[1] Originator of this idea is Gary Grutzek.

3.2.2 Spatial Sampling Strategies

The question of finding an appropriate distribution of sampling nodes on the sphere is a non-trivial mathematical problem. The target of designing suitable sampling schemes is the stable and unique identification of spherical harmonics up to a certain order, based on the sparse information that is acquired at as few sampling nodes as possible. A unique identification of the spherical harmonic modes is crucial in order to maintain orthogonality, cf. Eq. (2.85). Since the structure of the spherical harmonics gets progressively more complex with rising orders, any sampling distribution involving a limited number of nodes reaches a certain limit of modal resolution. Discrete sampling schemes resolve spherical harmonics up to a dedicated maximum order N_{sg} only. For strictly band-limited functions on the sphere ideal conditions can be assumed, while neglecting some minor approximation errors coming with certain types of sampling schemes. Unfortunately, natural sound fields are not order-limited. As a consequence, ambiguities concerning the mode identification arise, which result in spatial aliasing artifacts, cf. Section 3.8.

Spatial sampling schemes on the sphere can be generated using different approaches. A comprehensive overview of several sampling approaches that are relevant in the present context, including an analysis of their particular properties, can be found in (Zotter, 2009a, pp. 69–82) and (Zotter, 2009b). In this thesis, we restrict to quadrature (or cubature) approaches that are commonly applied for this purpose. Quadratures allow for a straight-forward application of the discrete spatial Fourier transform, given in Eq. (3.3). Quadratures usually define spatial node positions $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})$ and corresponding quadrature weights $w_{g_{\mathrm{sg}}}$, where $g_{\mathrm{sg}} = [1, M_{\mathrm{sg}}]$ describes an integer index of the sampling positions and M_{sg} the total number of sampling nodes (equivalent to microphones or microphone positions here).

Throughout this work we assume any quadrature weights $w_{g_{\mathrm{sg}}}$ to be unnormalized yielding

$$\sum_{g_{\mathrm{sg}}=1}^{M_{\mathrm{sg}}} w_{g_{\mathrm{sg}}} = 4\pi. \tag{3.1}$$

A theoretical minimum boundary for the required number of nodes M_{sg} according to a grid order N_{sg} can be estimated using

$$M_{\mathrm{sg}} \approx \eta_{\mathrm{g}} (N_{\mathrm{sg}} + 1)^2, \tag{3.2}$$

where η_{g} depends on the specific sampling scheme and indicates the degree of overdeterminacy. $\eta_{\mathrm{g}} = 1$ represents the best possible grid efficiency, $\eta_{\mathrm{g}} > 1$ indicates a certain inefficiency. Very few grids, such as hyperinterpolation (Zotter, 2009b) or quadratures

proposed by Fliege and Maier (1999), achieve a theoretical $\eta_{\mathrm{g}} = 1$. Throughout this work, two different quadratures are used; equiangular Gauss quadratures (Stroud and Secrest, 1966) with $\eta_{\mathrm{g}} = 2$, as well as quite efficient equidistant Lebedev quadratures (Lebedev, 1977) with $\eta_{\mathrm{g}} = 1.3$, cf. Figure 3.1. Lebedev quadratures are defined for a limited subset of orders only. Both quadrature types turn out to be stable and both have certain advantages depending on the respective application.

Drawing a first important conclusion, discrete spatial sampling reduces the modal resolution of the system.

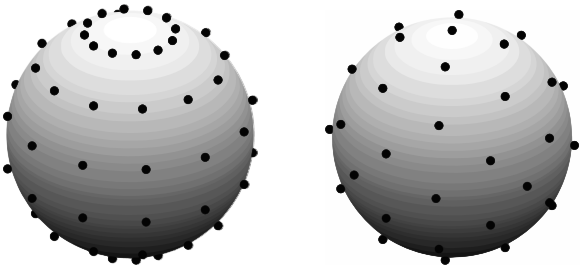


Figure 3.1 Equiangular Gauss quadrature with 72 nodes (left) and equidistant Lebedev quadrature with 50 nodes (right). Both have grid order $N_{\mathrm{sg}} = 5$. The reason for the difference in efficiency can be easily observed in the figure in this case; the Gauss quadrature provides disproportionately many nodes at the poles, while the lowest node density (around the equator) is the decisive factor. The Levedev quadrature provides nearly equally spaced nodes on the sphere, which is the reason for its higher efficiency.

3.2.3 Discrete Spatial Fourier Transform

Instead of having continuous knowledge of the pressure on the surface S_0, we only have information at discrete quadrature nodes $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}}) \in S_0$ at this point. Hence, the continuous spatial Fourier transform given in Eq. (2.92) is not applicable. The corresponding discrete spatial Fourier transform for the estimation of spatial Fourier coefficients yields

$$\mathring{G}_{nm}(r_0, \omega) \approx \sum_{g_{\mathrm{sg}}=1}^{M_{\mathrm{sg}}} w_{g_{\mathrm{sg}}}\, G(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}}, r_0, \omega)\, Y_n^m(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})^*, \tag{3.3}$$

where $w_{g_{\mathrm{sg}}}$ denotes the quadrature weights. Note that only for $n \leq N_{\mathrm{sg}}$ valid spatial Fourier coefficients can be deduced. As we use quadrature based sampling schemes here, approximation errors are expected. However, the errors are negligible in practice.

3.3 PWD and Modal Beamforming with Limited Modal Order

Discrete spatial sampling reduces the maximum resolvable modal order of the system. Limiting the system's resolution to a maximum order $N\,(\leq N_{\mathrm{sg}})$ has direct impact on the properties of the PWD and modal beamforming.

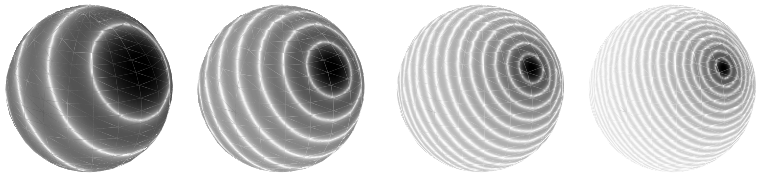
The order-limited PWD in contrast to the ideal PWD from Eq. (2.143) yields

$$D(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c}\, r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}). \tag{3.4}$$

Analogous, order-limited modal beamforming yields

$$B'(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{\mathring{B}_{nm}^{(\mathcal{R})}(\omega)}{\mathrm{i}^n \, j_n(\frac{\omega}{c}\, r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}). \tag{3.5}$$

Both expressions are quite similar to their ideal counterparts from Section 2.19, except for the first sum running only up to the maximum finite order N instead of $N \to \infty$.



N=5 N=10 N=20 N=30

$0 \longleftarrow$ no spatial resolution (omni-directional)

perfect spatial resolution $\longrightarrow \infty$ (spatial Dirac pulse)

Figure 3.2 Normalized logarithmic magnitude of the PWD for a single plane wave impact at different exemplary orders $N = \{5, 10, 20, 30\}$ projected to a spherical surface with a limited dynamic range between -50 dB (white) and 0 dB (black) for better visualization. The corresponding magnitudes are shown in Figure 3.3.

The order reduction has a severe impact on both operations. The ideal PWD for $N \to \infty$ describes a spatial dirac pulse with an infinitely small beam (infinite directional gain) towards a target direction and perfect rejection of any other direction. Truncation of the order N entails a widened main-lobe (which means lower spatial resolution) and the appearance of additional side-lobes. The effects of truncating the modal order are illustrated in Figure 3.2 and Figure 3.3. Towards infinite orders $N \to \infty$, the response approaches a spatial Dirac pulse with perfect spatial resolution. In contrast,

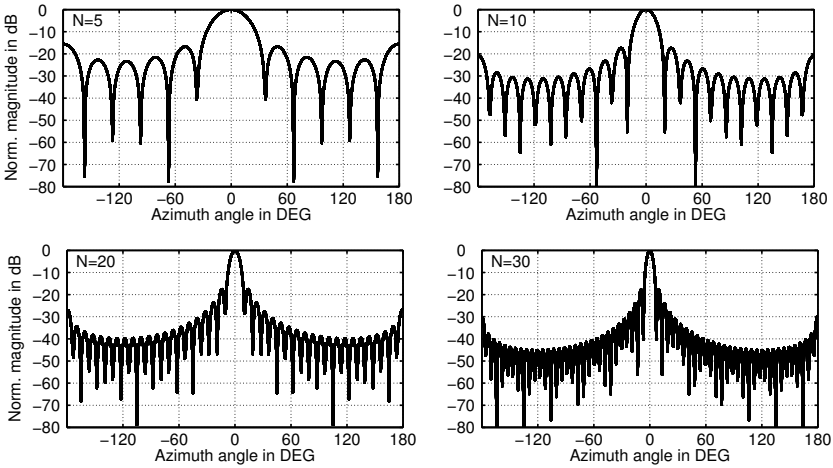


Figure 3.3 Normalized magnitude response of the PWD (regular beampattern) in the azimuth plane ($\theta_\mathrm{d} = \pi/2$, $\phi_\mathrm{d} = [-\pi, \pi]$) for a plane wave impact from ($\theta_\mathrm{w} = \pi/2$, $\phi_\mathrm{w} = 0$) at modal orders $N = \{5, 10, 20, 30\}$.

for $N = 0$ no spatial resolution can be achieved, since we only use the first omni-directional spherical harmonic mode. The total number of lobes including main-lobe, back-lobe, and all side-lobes is $N + 1$. Low truncation orders bring up few but strong side-lobes, high orders bring up more but weaker side lobes. For $N \to \infty$ the side-lobes vanish.

At this point, the relation of the PWD and classical beamforming approaches known from arbitrary sensor arrays (e.g. antenna arrays) becomes apparent. The order-truncated PWD with $\mathring{B}_{nm} = 1$ is commonly referred to as regular beampattern. Classical beamforming objectives, such as controlling the directional gain or maintaining a minimum side-lobe rejection level etc., can be established using the beamforming coefficients $\mathring{B}_{nm}(\omega)$. Approaches for applying these classical beamforming objectives known from communications system theory in the modal domain can be found e.g. in (Teutsch, 2007), (Koretz and Rafaely, 2009), or (Agmon et al., 2009). These topics are not further discussed in the following, since all presented techniques are based on PWD with regular beampattern.

An important measure for beamformers, as well as for directional microphones, is the directivity index (DI), which describes the maximum achievable directional gain

referred to an omni-directional receiver. The DI for modal beamformers can be simply calculated using (Meyer and Elko, 2002)

$$\mathrm{DI} = 10 \log \left[(N+1)^2 \right]. \tag{3.6}$$

The DI of the modal beamformer for orders $N = [0, 30]$ is plotted in Figure 3.4. For a first order beamformer ($N = 1$), the relation of modal beamforming and classical directional microphones becomes visible. The hypercardioid microphone is a first order transducer with maximum directional gain. Both, the modal beamformer with $N = 1$ and the hypercardioid microphone, have a directivity index $\mathrm{DI} = 6.02\,\mathrm{dB}$.

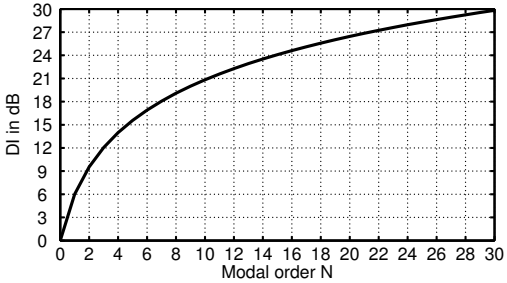


Figure 3.4 Directivity index (DI) of a modal beamformer with $N = [0, 30]$.

A remarkable property of modal beamforming is the native frequency-independence of the beam, which is commonly referred to as constant directivity (CD) beamforming. Of course, the beam can be intentionally modified by setting frequency-dependent coefficients $\mathring{B}_{nm}(\omega)$, but while applying frequency-independent beamforming coefficients, the beampattern is frequency-independent likewise. This property is owed to the frequency-independence of the spherical harmonics. The CD property is illustrated in Figure 3.5. Unfortunately, constraints in technical systems usually impede maintaining true CD behavior on the entire audio spectrum in practice.

3.4 Composite Signal

It might appear logically consistent that integrating ideal PWD signals (i.e. a continuous distribution of ideal spatial Dirac pulses) over the closed sphere S using

$$C'(\omega) = \frac{1}{4\pi} \iint_S \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c} r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta, \phi) \, \mathrm{d}\Omega \tag{3.7}$$

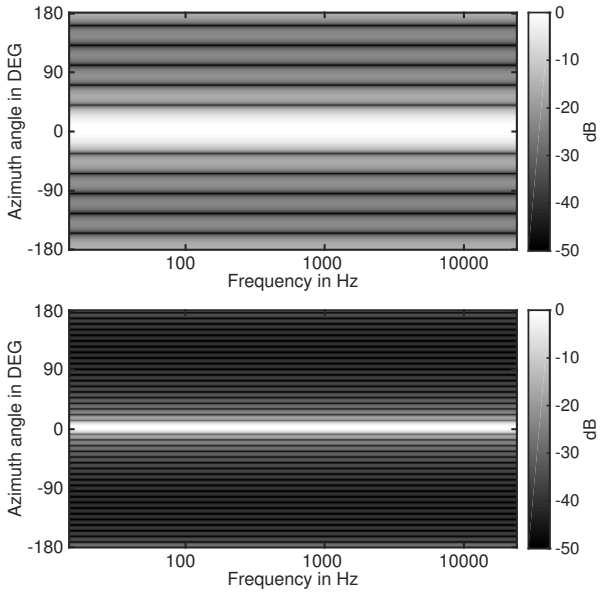


Figure 3.5 Normalized magnitude response $|D(\theta_\mathrm{d}, \phi_\mathrm{d}, \omega)/(N+1)^2|$ of the PWD (regular beampattern) in the azimuth plane ($\theta_\mathrm{d} = \pi/2$, $\phi_\mathrm{d} = [-\pi, \pi]$) for an order-limited broadband plane wave impact from ($\theta_\mathrm{w} = \pi/2$, $\phi_\mathrm{w} = 0$) at decomposition orders $N = 5$ (top) and $N = 20$ (bottom) versus the temporal frequency. The beampattern is constant. Note that the order of the wave is matched to the order of the decomposition here.

yields an omni-directional output signal $C'(\omega)$ that exactly corresponds to the signal $C_0(\omega)$ that can be measured using an ideal omni-directional transducer at the origin of the sphere S_0. Hence, we find the relation $C'(\omega) = C_0(\omega)$.

However, in practice, we face two limitations concerning this equation. Since we assume the order of the source system delivering $P_{nm}(r_0, \omega)$ to be limited to any finite order N, the order of the first sum needs to be limited to N.

Moreover, continuous integration over the surface S cannot be performed in practice. Hence, we must restrict ourselves to a finite number of discrete spatial nodes on S in order to be able to perform numeric integration in this context. It appears reasonable to fall back on quadratures (cf. Section 3.2.2), since they are precisely designed for performing numerical integration on the sphere. In addition, quadratures provide a valid space-frequency domain base for spherical harmonics up to the order of the

quadrature. Thus, the continuous integration over S is exchanged with a summation over the discrete quadrature nodes $(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}})$, $g_{\mathrm{cg}} = [1, M_{\mathrm{cg}}]$, with M_{cg} denoting the total number of nodes. The nodes $(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}})$ need to be scaled with individual quadrature weights $w_{g_{\mathrm{cg}}}$ to support equality, cf. Section 3.2.2. Finally, the entire expression reads

$$C(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}} \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c} \, r_0)} \, \mathring{P}_{nm}(r_0, \omega) \, Y_n^m(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}). \tag{3.8}$$

The involved quadrature is referred to as composite grid with grid order N_{cg} in the following, as it is used to recompose the previously decomposed wave field. The signal $C(\omega)$ is referred to as composite signal, accordingly. Even under the given constraints we still find $C(\omega) = C_0(\omega)$, while the composite grid order fulfills $N_{\mathrm{cg}} \geq N$. This once more points to the closed form of the spherical harmonic decomposition approach and the suitability of quadratures serving as discrete space-frequency domain base. Eq (3.8) and its logical consequences, i.e. $C(\omega) = C_0(\omega)$ for $N_{\mathrm{cg}} \geq N$ are of great importance in the following.

3.5 Binaural Systems with Limited Modal Resolution

Limited modal resolution is a fundamental problem for binaural reproduction. Modal sound field descriptions that are truncated at low orders turn out to be incompatible with HRTFs. This issue is discussed in (Bernschütz, 2014) and (Bernschütz et al., 2014).

In order to point out the elementary problem, we begin with analyzing the modal properties of HRTFs. As discussed in Section 2.20, a spherical HRTF set $H^{\mathrm{l,r}}(\theta, \phi, \omega)$ can be regarded as a complex function of magnitude and phase on the sphere. Hence, the spatial Fourier transform is applicable in order to obtain spatial Fourier coefficients $\mathring{H}^{\mathrm{l,r}}_{nm}(\omega)$. Figure 3.6 shows the modal intensity distribution of a measured HRTF set that is described in Section 4.3.2.

As is clearly visible from Figure 3.6, the HRTFs progressively involve higher modes at higher temporal frequencies. The higher orders arise due to air-path delays and complex physical scattering effects around the head. Both are of fundamental importance for establishing localization cues that can be read by the human auditory system for the localization of sound sources. Thus, in order to resolve HRTFs that cover the entire audible time-frequency bandwidth and that are capable of establishing the necessary localization cues for binaural auralization, the system should contain modal orders[2] up to approximately $N = 35$.

[2] The given value refers to HRTFs without torso, cf. Section 4.3.2.

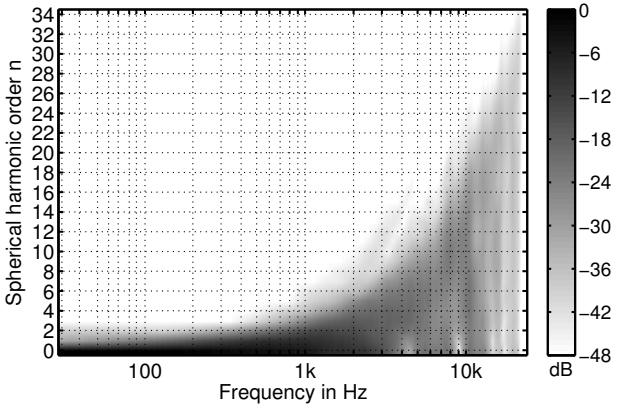


Figure 3.6 Modal intensity distribution for one ear channel of a measured spherical HRTF set versus the temporal frequency. Contributions of all modes $m = [-n, n]$ in a specific order n are summed. The data set is normalized to 0 dB and the dynamic range is limited to −48 dB for better visualization.

With a sampling scheme with highest possible efficiency ($\eta_\mathrm{g} = 1$), an array with around 1300 nodes (microphones) would be required to strictly fulfill this claim. This is obviously not feasible in practice, unless when using a scanning microphone array. The latter, however, is not capable of performing real-time operation. There are also other technical constraints that decrease the maximum achievable modal bandwidth of microphone arrays, which are discussed during the next sections. A reasonable modal resolution of microphone arrays is in the range of approximately $N = [4, 7]$, far away from $N = 35$. A fundamental adaptation problem arises, which immediately becomes apparent in the modal domain. The problem is illustrated in Figure 3.7.

The source system (microphone array) only delivers lower orders $N = [0, 5]$ in the given example. There is no information available to feed the higher modal inputs of the HRTF set. Setting the residual inputs to zero, which means truncating the order of the HRTF set, would entail severe loss of information at higher temporal frequencies. Besides the inherent loss of spatial resolution in the modal domain, we expect some kind of low-pass characteristics to arise in the time-frequency domain, since the signal contributions coded in higher modes would get lost without substitution.

The effects of the order-truncation can be be evaluated applying an order-limited inverse spatial Fourier transform to the spatial Fourier coefficients $\mathring{H}^{\mathrm{l,r}}_{nm}(\omega)$ that describe

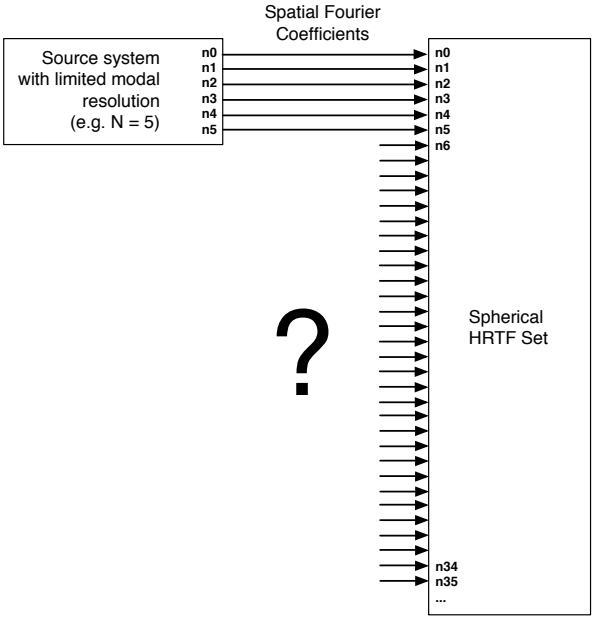


Figure 3.7 Illustration of the adaptation problem between a source system with limited modal resolution and a spherical HRTF set in the modal domain.

the spherical HRTF set in the modal domain. The inverse spatial Fourier transform is introduced in Eq. (2.96) and the order-limited inverse transform of the HRTFs yields

$$H_N^{\mathrm{l,r}}(\theta, \phi, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \mathring{H}_{nm}^{\mathrm{l,r}}(\omega)\, Y_n^m(\theta, \phi), \tag{3.9}$$

where $H_N^{\mathrm{l,r}}(\theta, \phi, \omega)$ describes a HRTF in the space-frequency domain that is limited to the modal order N in the spherical wave spectrum domain. The angle pair (θ, ϕ) denotes the source position. Figure 3.8 shows the impact of the order-truncation in the modal domain on the spectral magnitude in the time-frequency domain for two exemplary source positions or head rotations, i.e. a frontal and a lateral HRTF.

For the frontal HRTF, depicted in Figure 3.8 a), we can indeed observe the expected low-pass effect that decreases with increasing truncation order. The low-pass effect is not constant and depends on the source position or head rotation. If the ear is e.g. turned towards the source (lateral HRTF), the low-pass effect nearly vanishes and the HRTF appears to be comparatively well reproduced even at lower truncation orders.

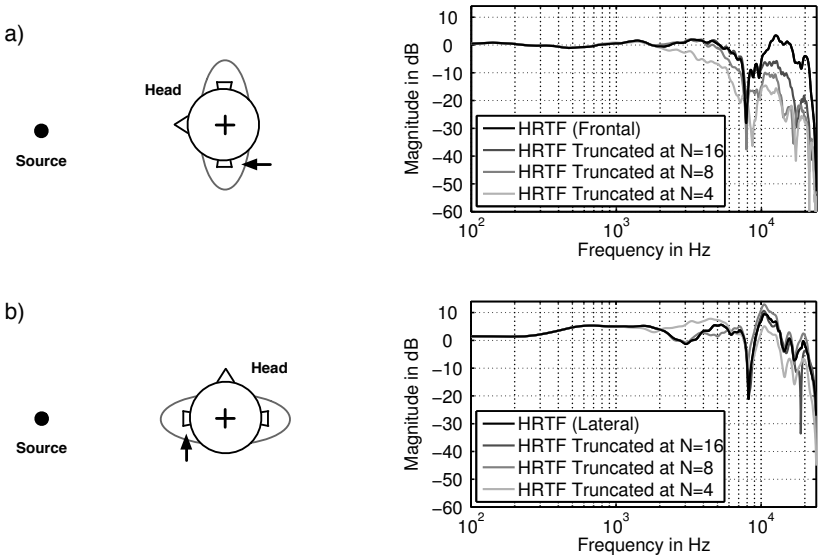


Figure 3.8 Impact of truncating the order of HRTFs in the modal domain on the spectral magnitude in the time-frequency domain for a) a frontal and b) a lateral HRTF. The left ear channel is evaluated.

A corresponding problem and identical phenomena must be observable in the space domain, since the modal domain and the space domain are unambiguously connected through the spatial Fourier transform.

In theory, a single far-field HRTF, $H^{\mathrm{l,r}}(\theta, \phi, \omega)$, in the space domain is assumed to be acquired using a point source located at infinite radius in the direction (θ, ϕ). Using ideal PWD, the spatial Dirac pulse in the space domain is capable of resolving this constellation, since it allows for perfect spatial sampling. For this reason, the ideal operation is claimed to be transparent in Section 2.20.

In contrast, considering PWD with limited modal order, we observe a certain degeneration of the spatial Dirac pulse that manifests itself in a widened main-lobe and the appearance of additional side-lobes. Neglecting the side-lobes for simplicity, an ideal point source would be interpreted as a source with dedicated spatial expansion according to the width of the main-lobe. The width of the main lobe increases as the truncation order of the PWD decreases. The previously assumed point source is now

translated into an isophasic expanded spherical source, which basically excites several HRTFs at once. Since the ear is not placed at the physical origin, which is the mathematical reference of the PWD, different propagation path lengths arise. Different path lengths entail different phase shifts. As a consequence, the overlay of the signals shows cancellations, especially at high frequencies, where the path lengths start to exceed half cycle of the wave length. This implies low-pass characteristics of the transmission system in the time-frequency domain.

To clarify this, a simplified monopole array analogy can be used as illustrated in Figure 3.9. Due to the order truncation of the PWD a dense array of monopole sources is located on a spherical cap that has an expansion according to the main lobe width of the degenerated Dirac pulse. The spherical cap is embedded in an imaginary sphere around the physical origin (mathematical reference of the PWD).

In Figure 3.9 a), the receiver is placed at the origin and the path lengths from all radiating monopoles are identical. Due to the phase coherence, the contributions always sum up perfectly, independent of the cap expansion. We obtain a flat magnitude response in the time-frequency domain. As soon as HRTFs are involved, the receiver is located with a certain spatial offset from the origin, cf. Figures 3.9 b) and c).

The path lengths are different and depend either on the head rotation or the direction of sound incidence. At low frequencies, the path differences are widely negligible but at higher frequencies with shorter wavelengths they evoke intense cancellations. This leads to low-pass effects that depend both on the order of the truncation, thus on the main lobe width, and on the head rotation or sound incidence direction. If the ear is turned towards the source, cf. Figure 3.9 b), the path differences are comparably low, so is the manifestation of the low pass effect. If, by contrast, the head looks towards the source or the source is located behind the head, the path differences become larger, and, as a consequence, the low pass effect gets more prominent, cf. Figure 3.9 c).

A very similar behavior can be observed when truncating the modal order of the HRTF set, cf. Figure 3.8. The simplified analogy does not deliver exactly the same results, since neither the side lobes nor the decrease of intensity in the main lobe away from the center are considered. This overemphasizes the cancellation effects. Nevertheless, the elementary problem becomes clear. Identical problems and effects are observable in both the spherical wave spectrum and the space-frequency domain, indicating a fundamental adaptation problem. The following sections focus on the adaptation of HRTFs to systems providing lower modal order. Nevertheless, the approach is generally applicable for the adaption of two arbitrary system with different modal resolution.

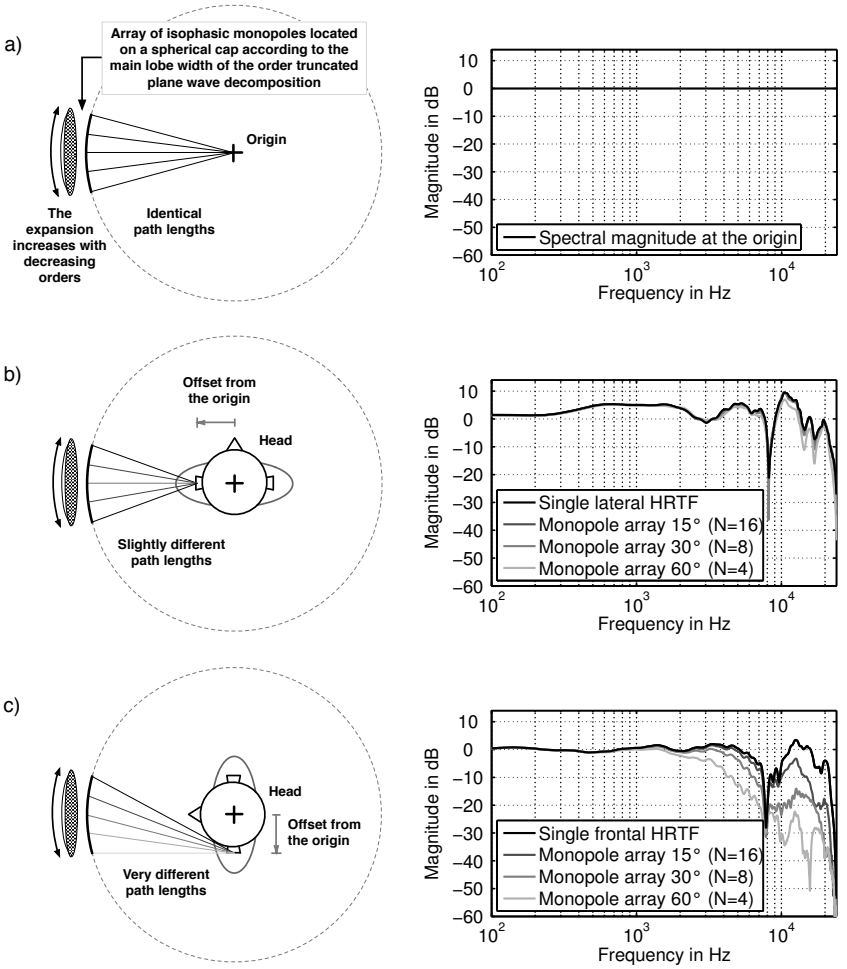


Figure 3.9 Simplified monopole array analogy in the space domain for pointing out the modal adaptation problem.

In order to find a suitable adaptation approach that conserves the spectral transmission properties in the space-frequency domain in the best possible way, we return to equation Eq. (2.145) that is used for generating the binaural output signal $Y^{\mathrm{l,r}}(\omega)$ from a modal sound field description $\mathring{P}_{nm}(r_0,\omega)$ under ideal conditions,

$$Y^{\mathrm{l,r}}(\omega) = \frac{1}{4\pi} \iint_S H^{\mathrm{l,r}}(\theta,\phi,\omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c}\, r_0)} \, \mathring{P}_{nm}(r_0,\omega) \, Y_n^m(\theta,\phi) \, \mathrm{d}\Omega.$$

We find the latter equation to be very similar to Eq. (3.7) that describes the ideal recomposition, except for the introduced HRTF weighting by $H^{\mathrm{l,r}}(\theta,\phi,\omega)$. In practice, we face the identical constraints concerning the finite order N of the source system delivering $P_{nm}(r_0,\omega)$ on the one hand and continuous integration over S on the other hand. The summation is limited to N and the continuous integration is replaced by a discrete composite grid (quadrature), analogous to the transition from Eq. (3.7) to Eq. (3.8),

$$Y^{\mathrm{l,r}}(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}} \, H^{\mathrm{l,r}}(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}},\omega) \times \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c}\, r_0)} \, \mathring{P}_{nm}(r_0,\omega) \, Y_n^m(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}}). \quad (3.10)$$

The HRTFs must exactly match the composite grid angles $(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}})$. Thus, either the involved HRTF set is measured with respect to matching composite grid angles $(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}})$ or the necessary signals are derived from a high-resolution HRTF set using spherical harmonic interpolation for spatial resampling,

$$Y^{\mathrm{l,r}}(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}} \left[\sum_{g'=1}^{M'} \sum_{n'=0}^{N'} \sum_{m'=-n'}^{n'} w'_{g'} \, H^{\mathrm{l,r}}(\theta_{g'},\phi_{g'},\omega) \, Y_n^m(\theta_{g'},\phi_{g'})^* \times Y_n^m(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}}) \right] \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{1}{\mathrm{i}^n \, j_n(\frac{\omega}{c}\, r_0)} \, \mathring{P}_{nm}(r_0,\omega) \, Y_n^m(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}}). \quad (3.11)$$

The spherical harmonic interpolation in square brackets consists of a forward spatial Fourier transform applied to the high-resolution HRTF set and a backward spatial Fourier transform to the required HRTF nodes $(\theta_{g_{\mathrm{sg}}},\phi_{g_{\mathrm{sg}}})$, which is discussed in greater detail in Section 3.12.4. The HRTFs need to be measured on a grid with M' nodes that is capable of resolving spherical harmonic orders $N' \geq 35$ for the HRTF set that

is used throughout this work and discussed in Section 4.3.2. The angles $(\theta_{g'}, \phi_{g'})$ refer to the measured nodes of the HRTF set and $w'_{g'}$ denotes the corresponding quadrature weights. While performing this operation, rotation of the HRTFs is performed using Wigner-D functions or applying respective angle offsets to $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})$ during the backward transform as previously discussed in Section 2.20.

So far, the major question remains, i.e. how to adapt the high-order HRTFs to the low order source system in a best possible way, getting rid of the inherent low-pass filter effects.

3.5.1 Spatial Subsampling of HRTFs

During the transition from Eq. (2.145) to Eq. (3.10), the continuous integration over the sphere S is replaced by a summation over discrete nodes of a quadrature grid that was previously defined as composite grid. The composite grid is the key for a suitable adaptation of the high-order HRTFs to a low-order source system, as will be shown in the following.

In order to achieve a best possible approximation to the continuous integration, it might appear reasonable to use a composite grid with the highest possible resolution, i.e. $M_{\mathrm{cg}} \to \infty$ ($N_{\mathrm{cg}} \to \infty$). At least, the composite grid might be expected to resolve the HRTF set properly, i.e. providing $N_{\mathrm{cg}} \geq 35$ in the given context. This means covering $M_{\mathrm{sg}} \geq 1730$ nodes using a Lebedev quadrature or $M_{\mathrm{sg}} \geq 2592$ nodes using a Gauss quadrature, for instance.

With a high-order composite grid, the HRTFs indeed unfold their full modal resolution, but as a consequence, we run into the modal adaptation problem discussed earlier.

However, considerable improvement of the adaptation can be achieved by using a low-order composite grid instead that exactly matches the order of the low-order source system, i.e. $N_{\mathrm{cg}} = N$. Thus, if the source system provides a maximum order of $N = 5$ for example, the composite grid should match $N_{\mathrm{cg}} = 5$ likewise. For the given example using $N_{\mathrm{cg}} = 5$, a Gauss composite grid consists of $M_{\mathrm{cg}} = 72$ and a Lebedev grid of $M_{\mathrm{cg}} = 50$ nodes only. Using a low-order composite grid indeed means to perform spatial subsampling of the HRTF set, since we are restricted to M_{cg} spatial nodes for sampling the entire spherical HRTF set. Using subsampling in the space-frequency domain enables us to properly reduce the modal resolution of the HRTF set and to avoid the spectral low-pass effects due to the modal mismatch of both systems.

3.5.2 HRTFs with Limited Modal Resolution (RHRTFs)

In order to demonstrate the adaptation enhancements using spatial HRTF subsampling, as well as to evaluate the impact of the modal reduction on HRTFs in general, suitable

signals need to be generated and appropriate measures need to be defined that allow for performing meaningful and informative analysis.

A reasonable reference for a comparative analysis is the original HRTF $H^{\mathrm{l,r}}(\theta,\phi,\omega)$. In contrast to the original HRTF, a processed signal is generated by inserting an analytic plane wave with incidence direction (θ,ϕ) defined by

$$\mathring{P}_{nm\,\mathrm{pw}(\theta,\phi)}(r_0,\omega) = 4\,\pi\,\mathrm{i}^n\,j_n(\tfrac{\omega}{c}\,r_0)\,Y_n^m(\theta,\phi)^* \tag{3.12}$$

into Eq. (3.11),

$$H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta,\phi,\omega) = \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}} \left[\sum_{g'=1}^{M'} \sum_{n'=0}^{N'} \sum_{m'=-n'}^{n'} w'_{g'}\, H^{\mathrm{l,r}}(\theta_{g'},\phi_{g'},\omega) \times Y_n^m(\theta_{g'},\phi_{g'})^*\, Y_n^m(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}}) \right] \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^m(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}})\, Y_n^m(\theta,\phi)^*. \tag{3.13}$$

The output signal $H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta,\phi,\omega)$ describes a HRTF with limited modal resolution with the plane wave order N and the composite grid order N_{cg}. For the subsequent simulations the HRTFs presented in Section 4.3.2 are used. The HRTF set $H^{\mathrm{l,r}}(\theta_{g'},\phi_{g'},\omega)$ is measured on a Lebedev grid with $M' = 2702$ spatial sampling nodes. Spatial resampling to the required composite grid nodes $(\theta_{g_{\mathrm{cg}}},\phi_{g_{\mathrm{cg}}})$ is inherently performed within Eq. (3.13) by employing spherical harmonic HRTF interpolation.

At this point we have two different signals $H^{\mathrm{l,r}}(\theta,\phi,\omega)$ and $H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta,\phi,\omega)$ that can be adequately compared. In the following, we compare the magnitudes $|H^{\mathrm{l,r}}(\theta,\phi,\omega)|$ and $|H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta,\phi,\omega)|$. Another option would be to evaluate the phase responses.

For evaluating several different angles (θ_j,ϕ_j), $j = [1,J]$ at once, it is more useful to restrain the data to a single value describing an appropriately averaged spectral deviation between $|H^{\mathrm{l,r}}(\theta_j,\phi_j,\omega)|$ and $|H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta_j,\phi_j,\omega)|$ for a specific angle (θ_j,ϕ_j). We introduce the measure $\Delta_{\mathrm{E}}(\theta_j,\phi_j,N,N_{\mathrm{cg}})$ for that purpose. To simplify matters, the calculation of the latter is described by the informal equation

$$\Delta_{\mathrm{E}}(\theta_j,\phi_j,N,N_{\mathrm{cg}}) = \left| 20\,\log_{10}\left(\mathrm{avg}\left[\overbrace{\mathrm{avg}_{1/3\,\mathrm{oct}}\left[\frac{|H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}(\theta,\phi,\omega)|}{|H^{\mathrm{l,r}}(\theta,\phi,\omega)|} \right]}^{1-20\,\mathrm{kHz}} \right] \right) \right|. \tag{3.14}$$

The magnitude deviation is calculated by division of $|H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}|$ by $|H^{\mathrm{l,r}}|$ in a first step. In a second step, the spectral values are grouped and pre-averaged in 1/3 oct bands

to account for logarithmic frequency scaling that relates to our perception. Deviations in the high frequency range would be overemphasized otherwise. The evaluation is restricted to a temporal frequency range of $1 - 20\,\mathrm{kHz}$, since the spectral deviation is negligible below $1\,\mathrm{kHz}$. In a third step, the resulting $1/3\,\mathrm{oct}$ block values are averaged and converted to decibels. In a fourth step, the absolute value of the decibel value is taken. Although it might not appear perfectly reasonable from a perceptual point of view, positive and negative deviations are treated equal, which simply allows for a clearer graphical representation of the data set.

In order to obtain an even more abstract measure that accounts for the mean spectral magnitude deviation over the entire sphere S for a certain wave order N and composite grid order N_{cg} in a single value, we define

$$\widehat{\Delta}_{\mathrm{E}}(N, N_{\mathrm{cg}}) = \frac{1}{J} \sum_{j=0}^{J} w_j \, \Delta_{\mathrm{E}}(\theta_j, \phi_j, N, N_{\mathrm{cg}}). \tag{3.15}$$

For the subsequent simulations an equidistant Lebedev quadrature with $J = 974$ nodes (θ_j, ϕ_j) and weights w_j is used to acquire representative average deviation values over the entire sphere S. $\widehat{\Delta}_{\mathrm{E}}(N, N_{\mathrm{cg}})$ for $N \in [3, 9] \times N_{\mathrm{cg}} \in [3, 14]$ is depicted in Figure 3.10 using two different composite grid types, i.e. Lebedev and Gauss quadratures, in order to evaluate the adaptation properties using different wave orders vs. composite grid orders.

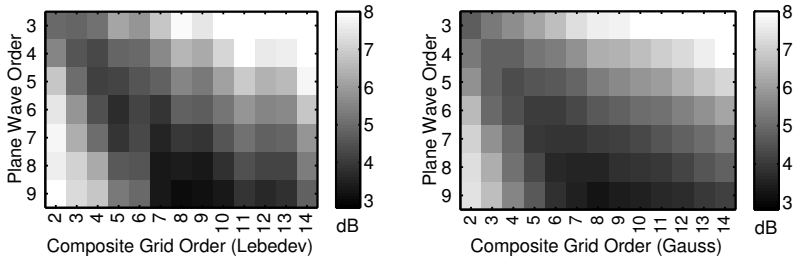


Figure 3.10 $\widehat{\Delta}_{\mathrm{E}}(N, N_{\mathrm{cg}})$ in dB coded in the gray scale for $N \in [3, 9] \times N_{\mathrm{cg}} \in [3, 14]$.

Figure 3.10 clearly shows the minimum of the magnitude deviation for $N_{\mathrm{cg}} = N$, which indicates that the generated signal $H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}$ is most similar to the original HRTF $H^{\mathrm{l,r}}$ for this constellation. Composite grids with $N_{\mathrm{cg}} > N$ or with $N_{\mathrm{cg}} < N$ lead to larger deviations of the generated signal. Hence, spatial subsampling of the HRTF set to the order of the source system indeed optimizes the overall transmission properties of the coupled system.

As previously discussed, the deviation decreases for increasing wave orders, since both signals are claimed to become identical for high orders, i.e. $H^{\mathrm{l,r}}_{(N,N_{\mathrm{cg}})\to\infty} = H^{\mathrm{l,r}}$. In practice, an order of $N_{\mathrm{cg}} = N \geq 35$ can be considered sufficient to achieve just about identical signals with negligible deviation on the entire audible spectrum, i.e. we assume $H^{\mathrm{l,r}}_{(N,N_{\mathrm{cg}})\geq 35} = H^{\mathrm{l,r}}$.

Furthermore, we observe certain differences between the Lebedev and the Gauss composite grid, which will turn out to be a crucial factor of influence later on.

Since we have shown that order-matched composite grids $N_{\mathrm{cg}} = N$ achieve the most convenient adaptation properties, we restrict the discussion to this condition hereafter. This implies a constant relation between the modal system order and the composite grid order, which makes the additional index N_{cg} dispensable. Hence, $H^{\mathrm{l,r}}_{N,N_{\mathrm{cg}}}$ from Eq. (3.13) is reduced to $H^{\mathrm{l,r}}_{N}$. Analogous, $\Delta_{\mathrm{E}}(\theta_j, \phi_j, N, [N_{\mathrm{cg}}])$ is reduced to $\Delta_{\mathrm{E}}(\theta_j, \phi_j, N)$. $H^{\mathrm{l,r}}_{N}(\theta, \phi, \omega)$ is referred to as head-related transfer function with limited modal resolution (RHRTF).

3.5.3 Properties of RHRTFs

We have found the RHRTF to be identical to the HRTF for high modal orders. Nevertheless, even though the adaptation using spatial subsampling reduces the deviations, the signals cannot be expected to be identical when scaling down the modal order of the RHRTF, which is shown in Figure 3.10. In order to gain more specific information on the properties of RHRTFs, $\Delta_{\mathrm{E}}(\theta_j, \phi_j, N)$ is evaluated using a low wave order of $N = 5$ and a Lebedev composite grid in steps of 1° over the entire sphere S, i.e. $\theta_j \in [0°, 180°] \times \phi_j \in [0°, 359°]$. The result is depicted in Figure 3.11 using a Mollweide map projection (Snyder, 1997), (Bernschütz, 2012b) for better illustration. The left ear signal is used for evaluation. The ear position is marked. For judging the overall transmission properties, we need to keep in mind that a RHRTF consists of two ear channels that must be considered as a coupled pair. Neglecting possible anatomic asymmetries, the opposite ear delivers equivalent results. The map just needs to be mirrored horizontally for the opposite ear.

The auditory system combines both ear signals. As the influence of two independently modified ear signals on the perception is unknown in the given context, a reasonable combined analysis of both ear signals is hardly possible. For this reason, we further restrict the analysis to single ear signals. Nevertheless, a suitable combined analysis is inherently performed during the listening experiments in Section 5.

The deviations vary considerably depending on the wave incidence direction. Waves that arrive at the opposite side of the ear tend to entail larger deviations. Figures 3.12 to 3.14 show exemplary HRTF and their corresponding RHRTF magnitudes.

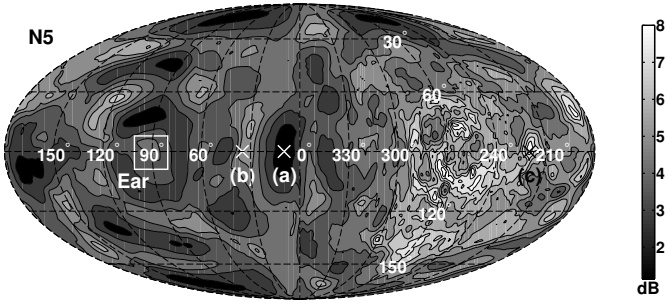


Figure 3.11 $\Delta_{\mathrm{E}}(\theta_j, \phi_j, N)$ (values encoded using different gray intensities) using a low wave order of $N = 5$ and a Lebedev composite grid for incidence directions $\theta_j \in [0°, 180°] \times \phi_j \in [0°, 359°]$. The markers (a), (b) and (c) refer to subsequent Figure 3.12, Figure 3.13 and Figure 3.14.

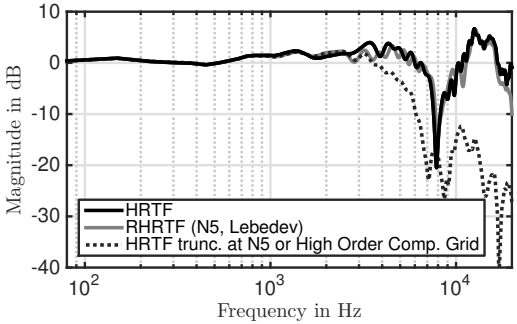


Figure 3.12 HRTF and RHRTF ($N = 5$, Lebedev) magnitudes for an incidence direction according to marker (a) in Figure 3.11. The deviation is rather small in this case. The dotted line shows the same HRTF when simply truncating its modal order to $N = 5$ or using a high order composite grid instead of order-matched subsampling.

Figure 3.12 shows a case with low deviation. Figure 3.13 is a random example that approximately corresponds to the overall mean deviation. Figure 3.14 shows a particularly bad example that corresponds to the maximum deviation for the entire RHRTF set. The dotted line in Figure 3.12 shows the result for the respective HRTF that is truncated at a modal order of $N = 5$ instead of using spatial subsampling. The identical magnitude response is achieved using a high order composite grid instead of a matched-order composite grid. The low-pass effect is clearly visible.

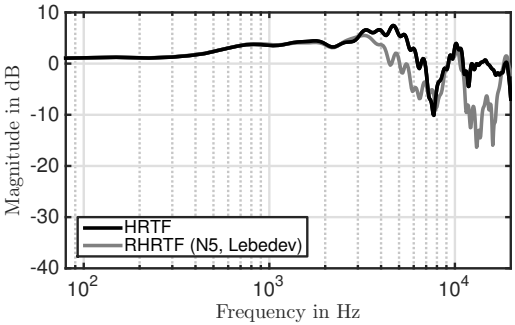


Figure 3.13 HRTF and RHRTF ($N = 5$, Lebedev) magnitudes for an incidence direction according to marker (b) in Figure 3.11. The deviation in this example approximately corresponds to the mean overall deviation.

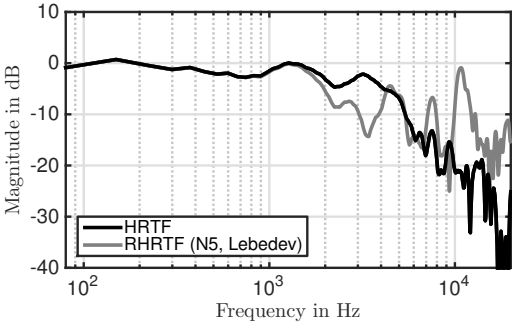


Figure 3.14 HRTF and RHRTF ($N = 5$, Lebedev) magnitudes for an incidence direction according to marker (c) in Figure 3.11.

Figure 3.15 is generated analogous to Figure 3.11 but using Gauss composite grids instead of a Lebedev grid. Comparing Figure 3.11 and the top image of Figure 3.15 reveals the influence of using different composite grid types. Different RHRTF orders $N = \{5, 7, 11\}$ are provided to illustrate the trend of Δ_{E} for increasing orders. As expected, the deviation Δ_{E} globally decreases at higher wave orders. The RHRTFs become more similar to the HRTFs for all incidence directions.

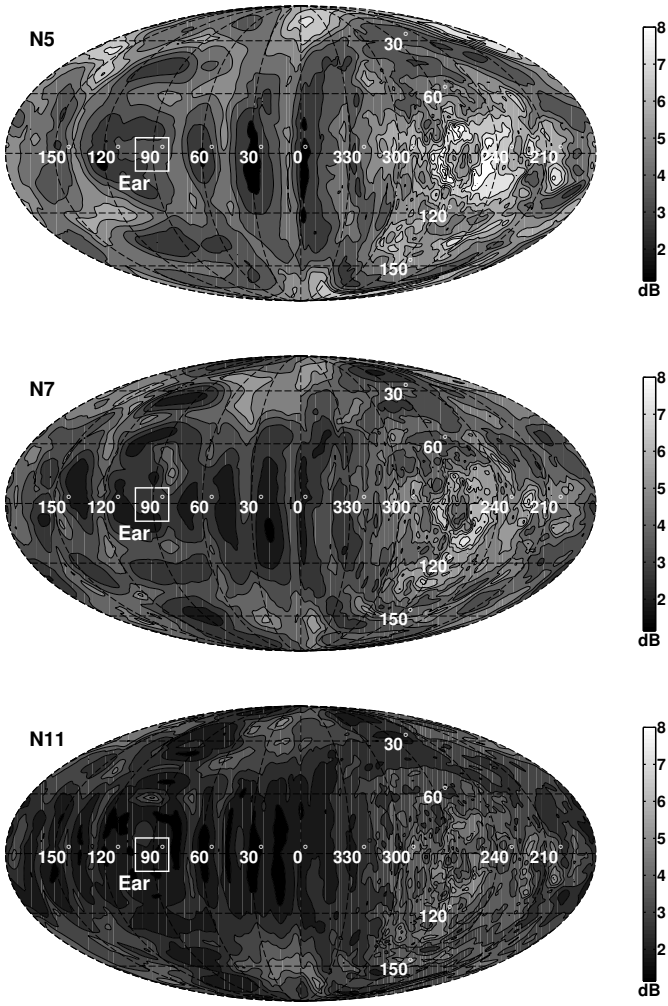


Figure 3.15 $\Delta_{\mathrm{E}}(\theta_j, \phi_j, N)$ (values encoded using different gray intensities) for different wave orders $N = \{5, 7, 11\}$ using Gauss composite grids.

Limited modal resolution and an appropriate adaptation of the HRTFs can clearly be regarded as an important factor of influence concerning the binaural reproduction of microphone array signals or modal sound field descriptions in general. So far this has not been discussed in the literature. Considerable deviations arise between HRTFs and low-order RHRTFs depending on the direction of incidence. However, a pure technical analysis or comparison of the ear signals does not allow for any reasonable assessment of the perceptual influences. Therefore, extensive listening experiments were performed, in order to assess the overall perceived audio quality of different RHRTFs compared to HRTFs on the one hand and specific perceptual attributes that are inherent to RHRTFs on the other hand. The listening experiments are presented in Section 5.

3.5.4 Positive Side-Effects of Modal Reduction and Spatial HRTF Subsampling

A convenient side-effect of the spatial subsampling approach is the dramatically reduced number of required HRTF nodes for building a system that is capable of representing sound incidence from arbitrary directions over the full sphere S and that enables free horizontal, vertical and lateral head rotation. Assuming e.g. a particularly low maximum system order of $N = 5$ using a Lebedev composite grid, we only need to measure $M_{\mathrm{cg}} = 50$ HRTF nodes distributed on the entire sphere.

The reduced number of nodes is highly convenient for individual HRTFs, cf. Section 3.12.1, since measuring full-sphere HRTF sets with high spatial resolution is time-consuming.

In addition, this approach provides considerable potential for designing scalable spatial audio data reduction algorithms.

3.6 Radial Filters

Radial filters are included in the PWD (cf. Section 2.19) and compensate for the radial portion of the Helmholtz equation in spherical coordinates that is solved by spherical Bessel and Hankel functions (cf. Section 2.10.2). The radial filters scale the amplification gain of the spherical harmonic modes. For each mode, a specific radial filter function is applied that depends on the measurement radius r_0 and the temporal frequency ω. The radial filters consist of $(N + 1)$ single time-frequency domain filters. In this work, the radial filters are implemented as finite impulse response (FIR) filters, which due to the non-causal portions adds some latency to the system. Alternative approaches for approximating infinite impulse response (IIR) radial filters are proposed e.g. in (Baumgartner et al., 2011).

3.6.1 Sphere Configuration

The radial filters depend on the sphere configuration. The sphere configuration describes whether the sensor nodes on S_0 are located in free field or are mounted on a rigid spherical body on the one hand, and the type of sensors (e.g. omni-directional, cardioid) (Rafaely, 2005) on the other hand.

Starting from Eq. (2.143) that describes the PWD and using a slightly different notation yields

$$D(\theta_\mathrm{d}, \phi_\mathrm{d}, \omega) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[\frac{1}{4\pi\, \mathrm{i}^n\, j_n(\frac{\omega}{c}\, r_0)} \right] \mathring{P}_{nm}(r_0, \omega)\, Y_n^m(\theta_\mathrm{d}, \phi_\mathrm{d}). \tag{3.16}$$

The term in square brackets describes the radial filters for omni-directional sensors located in free field (open sphere). In order to simplify the expression, the radial filters are expressed as (Rafaely, 2005)

$$d_n^{\mathrm{OS}}(\tfrac{\omega}{c}\, r_0) = \left[4\pi\, \mathrm{i}^n\, j_n(\tfrac{\omega}{c}\, r_0)\right]^{-1}. \tag{3.17}$$

The mode strength and the filter magnitudes for this configuration are depicted in Figure 3.16.

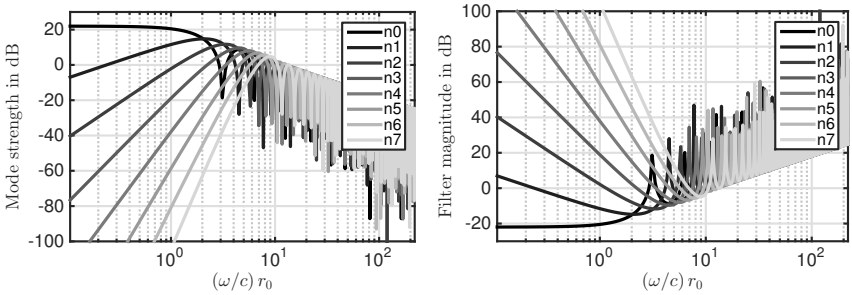


Figure 3.16 Mode strength (left) and magnitudes of the radial filters $|d_n^{\mathrm{OS}}(\frac{\omega}{c}\, r_0)|$ (right) for an open sphere array with omni-directional transducers at mode orders $n = [0, 7]$.

The open sphere configuration with omni-directional transducers is critical, since the mode strengths drops out and the filter amplification becomes infinite at the zeros of the spherical Bessel function in the denominator. The physical background is discussed in (Williams, 1999, pp 217–221). Hence, a single open sphere configuration with omni-directional transducers is not stable. Balmages and Rafaely (2007) propose a dual-

sphere approach to avoid this problem. The radii are chosen to complement the mode strength in the best possible way. The decomposer selects the more stable signal.

An alternative configuration is the open sphere array with cardioid transducers. The radial filters must account for the gradient portion of the cardioid transducers. They are defined by (Balmages and Rafaely, 2007)

$$d_n^{\mathrm{OSC}}(\tfrac{\omega}{c}\, r_0) = \left[4\,\pi\,\mathrm{i}^n\,\tfrac{1}{2}\,\left(j_n(\tfrac{\omega}{c}\, r_0) - \mathrm{i}\, j_n'(\tfrac{\omega}{c}\, r_0)\right)\right]^{-1}. \tag{3.18}$$

The mode strength and the filter magnitudes for this configuration are shown in Figure 3.17.

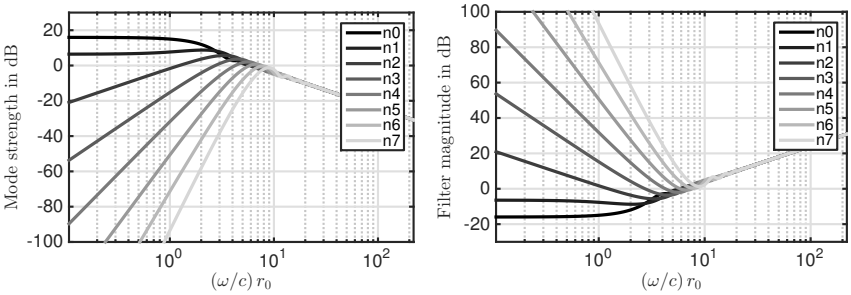


Figure 3.17 Mode strength (left) and magnitudes of the radial filters $|d_n^{\mathrm{OSC}}(\frac{\omega}{c}\, r_0)|$ (right) for an open sphere array with cardioid transducers at mode orders $n = [0, 7]$.

The minima of the mode strength can be avoided due to the gradient portion of the cardioid transducer. This configuration is suitable for practical application. However, cardioid microphones do usually not provide ideal cardioid properties in the entire temporal frequency range, which might entail impairments during the processing.

A common configuration is the rigid sphere array with omni-directional transducers. The tranducers are located on an acoustically hard, rigid spherical surface S_{RS} with radius $r_{\mathrm{RS}} \leq r_0$. The radial filters must account for the scattered field around the rigid body. The radial filters for this configuration yield (Meyer and Elko, 2002)

$$d_n^{\mathrm{RS}}(\tfrac{\omega}{c}, r_0, r_{\mathrm{RS}}) = \left[4\,\pi\,\mathrm{i}^n\left[j_n(\tfrac{\omega}{c}\, r_0) - \frac{j_n'(\frac{\omega}{c}\, r_{\mathrm{RS}})}{h_n'^{(2)}(\frac{\omega}{c}\, r_{\mathrm{RS}})} h_n^{(2)}(\tfrac{\omega}{c}\, r_0)\right]\right]^{-1}. \tag{3.19}$$

In most cases, the sensors are flush with the sphere, which implies $r_{\mathrm{RS}} = r_0$ and yields a simplified radial filter expression

$$d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0) = \left[4\,\pi\,\mathrm{i}^n \left[j_n(\tfrac{\omega}{c}\, r_0) - \frac{j_n'(\tfrac{\omega}{c}\, r_0)}{h_n'^{(2)}(\tfrac{\omega}{c}\, r_0)} h_n^{(2)}(\tfrac{\omega}{c}\, r_0)\right]\right]^{-1}. \tag{3.20}$$

The mode strength and the filter magnitudes for this configuration are depicted in Figure 3.18. The sound velocity is forced to zero at the rigid surface and hence the sound field is described by the sound pressure only. No dropouts in mode strength appear. The rigid sphere configuration turns out to be very stable in practice and is used throughout this work for simulations and measured data sets.

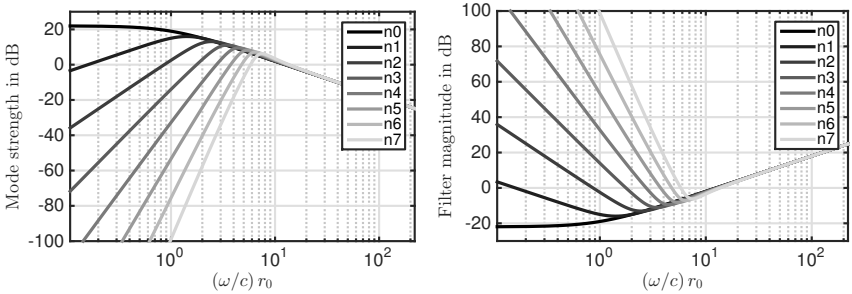


Figure 3.18 Mode strength (left) and magnitudes of the radial filters $|d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0)|$ (right) for a rigid sphere array with flush mounted omni-directional transducers at mode orders $n = [0, 7]$.

A disadvantage is the bulky appearance and difficult construction of rigid sphere arrays with large measurement radii. Furthermore, the array body cannot be assumed to be acoustically transparent. Even if the primary scattered field is taken into account by the radial filters, the array tends to reflect sound that might be thrown back by the environment, as long as the array is not placed in the free field. A rigid sphere array influences the reflection properties of the measured environment. However, for scenarios in the far-field using common array sizes (e.g. $r_0 \approx 0.1m$) this effect can be neglected.

Additional sphere configurations can be derived, but we restrict the discussion to the latter common configurations in this work.

With these filter definitions the PWD from Eq. (2.143) can be written as

$$D(\theta_\mathrm{d}, \phi_\mathrm{d}, \omega) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n^{\mathrm{OS/OSC/RS}}(\tfrac{\omega}{c}\, r_0)\, \mathring{P}_{nm}(r_0, \omega)\, Y_n^m(\theta_\mathrm{d}, \phi_\mathrm{d}). \tag{3.21}$$

This allows to account for different sphere configurations and provides a compact expression. For the order-limited PWD an analogous notation is applied.

In order to perform analysis based on simulated plane waves, analytic plane wave expressions are introduced that account for the sphere configurations given above.

For an open sphere array with omni-directional transducers the analytic plane wave is described by

$$\mathring{P}^{\mathrm{OS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w}, \phi_\mathrm{w})}(r_0, \omega) = 4\,\pi\,\mathrm{i}^n\, j_n(\tfrac{\omega}{c}\, r_0)\, Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^*. \tag{3.22}$$

The open sphere array with cardioid transducers yields

$$\mathring{P}^{\mathrm{OSC}}_{nm\,\mathrm{pw}(\theta_\mathrm{w}, \phi_\mathrm{w})}(r_0, \omega) = 4\,\pi\,\mathrm{i}^n\, \tfrac{1}{2} \left(j_n(\tfrac{\omega}{c}\, r_0) - \mathrm{i}\, j_n'(\tfrac{\omega}{c}\, r_0) \right) Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^*. \tag{3.23}$$

The rigid sphere array with omni-directional transducers, narrowed down to flush-mounted transducers (i.e. $r_\mathrm{RS} = r_0$), results in

$$\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w}, \phi_\mathrm{w})}(r_0, \omega) = 4\,\pi\,\mathrm{i}^n \left[j_n(\tfrac{\omega}{c}\, r_0) - \frac{j_n'(\frac{\omega}{c}\, r_0)}{h_n'^{(2)}(\frac{\omega}{c}\, r_0)} h_n^{(2)}(\tfrac{\omega}{c}\, r_0) \right] Y_n^m(\theta_\mathrm{w}, \phi_\mathrm{w})^*. \tag{3.24}$$

3.6.2 Example Configuration and Phase Responses

For further analysis, we define an exemplary rigid sphere array with $r_0 = 0.1\,\mathrm{m}$. Figure 3.19 shows the mode strength and radial filter magnitudes versus an absolute frequency scale for this configuration.

Besides of the magnitude responses, the particular phase responses of the complex-valued radial filters are of great interest. If the radial filters are to be modified in any way, it is of fundamental importance to maintain the relative phase relationships between the individual filters. Any individual processing using IIR filters is quite delicate. This is the main reason why falling back to FIR filters is preferable, as FIR filters allow for changing the magnitude response while maintaining the original phase response or to keep the phase relationship constant. The phase responses and the derived group delays are shown in Figure 3.20.

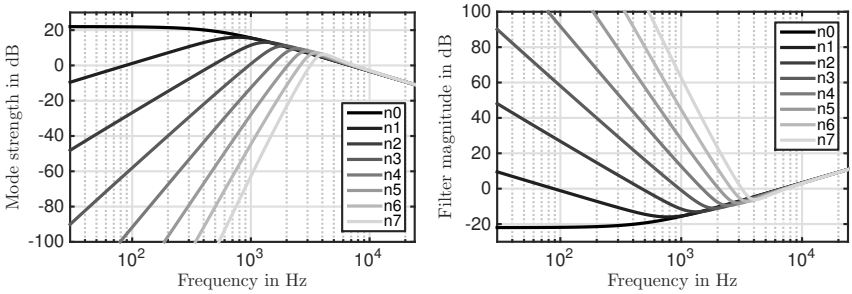


Figure 3.19 Mode strength (left) and magnitudes of the radial filters $|d_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ (right) for a rigid sphere array with omni-directional transducers and measurement radius $r_0 = 0.1\,\mathrm{m}$ at mode orders $n = [0, 7]$.

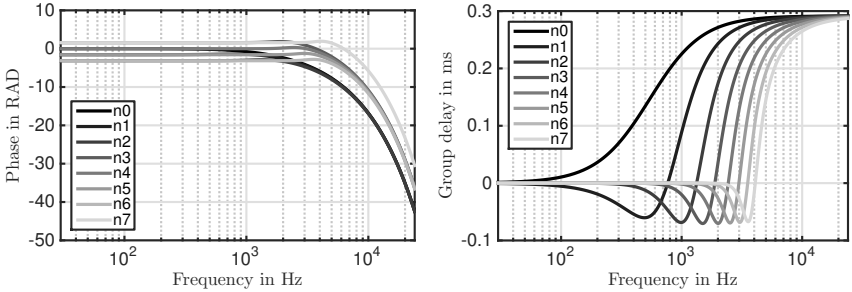


Figure 3.20 Phase response (left) and group delay of the radial filters $|d_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ (right) for a rigid sphere array with omni-directional transducers and measurement radius $r_0 = 0.1\,\mathrm{m}$ at mode orders $n = [0, 7]$.

3.6.3 Amplification Demands

One of the major issues concerning the radial filters are the excessive amplification demands for higher modal orders at low $(\omega/c)\, r_0$ ratios that can be observed in Figures 3.16 to 3.18. Figure 3.19 illustrates the amplification demands for a realistic array configuration.

In order to maintain the CD properties of the order-limited PWD over the entire temporal frequency range, illustrated in Figure 3.5, we need to apply the radial filters depicted in the right plot of Figure 3.19, since the mode strength at high orders and low temporal frequencies is weak, which is visible in the left plot of Figure 3.19.

The radial filters are critical in practice. Equipment noise from microphones, microphone amplifiers, and ADCs arises. External electromagnetic interference to the equipment or rounding noise on the digital stage can be considered as further but less relevant noise sources. The higher-order radial filters require amplification levels that quickly reach several hundreds of decibels at comparably low temporal frequencies. Thereby, equipment noise is amplified disproportionately compared to the useful signal, which can lead to a completely useless array response (Bernschütz et al., 2011b).

3.6.4 Limiting the Radial Filter Gain

The gain of the radial filters needs to be limited for achieving stable array responses in practice. However, setting a hard limit to the amplification leads to unsteady filter functions causing leaps in the spherical wave spectrum domain (Bernschütz et al., 2011b) and inconvenient time domain behavior of the system (Rettberg and Spors, 2014). In order to improve the system properties, a soft-knee limiting approach for limiting the amplifications gain of the radial filters is presented in (Bernschütz et al., 2011b). An arctangent function is employed for realizing the soft-knee characteristics yielding

$$\bar{d}_n(\tfrac{\omega}{c}\, r_0) = \frac{2\,\hat{a}}{\pi}\, \frac{d_n(\tfrac{\omega}{c}\, r_0)}{|d_n(\tfrac{\omega}{c}\, r_0)|}\, \arctan\left(\frac{\pi}{2\,\hat{a}} |d_n(\tfrac{\omega}{c}\, r_0)|\right), \tag{3.25}$$

where $\hat{a}$ denotes the linear amplification limit. Figure 3.21 (left plot) shows the radial filters $d_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$ from Figure 3.19 and Figure 3.21 (right plot) shows the identical filters with applied soft-knee limiting according to Eq. (3.25) yielding $\bar{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$. The amplification limit is set to $\hat{a}_{\mathrm{dB}} = 40\,\mathrm{dB}$ in the given example.

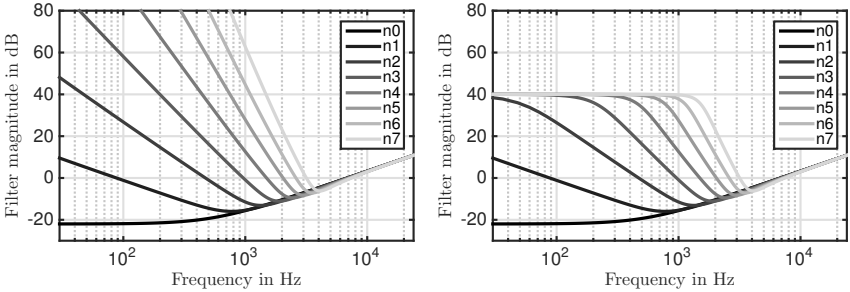


Figure 3.21 Radial filter magnitudes $|d_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ (left) and $|\bar{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ with an amplification limit of $\hat{a}_{\mathrm{dB}} = 40\,\mathrm{dB}$ (right).

For applying the gain limitation, the magnitudes of the filters need to be modified individually. Since the radial filters are applied as FIR filters, the magnitudes are modified while the original phase responses (see Figure 3.20) are left untouched.

Limiting the gain of the radial filters has an impact on the array response. Since the contribution of higher modal orders is reduced, the overall order of the array is reduced successively for decreasing temporal frequencies. As a consequence, the CD properties of the PWD cannot hold.

To illustrate this, we trace back to Eq. (3.24) that describes an ideal plane wave:

$$\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega) = 4\,\pi\,\mathrm{i}^n \left[j_n(\tfrac{\omega}{c}\,r_0) - \frac{j_n'(\tfrac{\omega}{c}\,r_0)}{h_n'^{(2)}(\tfrac{\omega}{c}\,r_0)} h_n^{(2)}(\tfrac{\omega}{c}\,r_0) \right] Y_n^m(\theta_\mathrm{w},\phi_\mathrm{w})^*.$$

Both $\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega)$ and $\bar{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\,r_0)$ are inserted into an order-limited PWD:

$$D^{\mathrm{L}}(\theta_\mathrm{d},\phi_\mathrm{d},\omega) = 4\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} \bar{d}_n^{\mathrm{RS}}(\tfrac{\omega}{c}\,r_0)\, \mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega)\, Y_n^m(\theta_\mathrm{d},\phi_\mathrm{d}). \quad (3.26)$$

The amplification limit is set to $\hat{a}_{\mathrm{dB}} = 40\,\mathrm{dB}$, cf. Figure 3.21 (right plot). An ideal plane wave from ($\theta_\mathrm{w} = \pi/2$, $\phi_\mathrm{w} = 0$) is generated using Eq. (3.24) and a scanning PWD of order $N = 7$ along the horizontal plane ($\theta_\mathrm{d} = \pi/2$, $\phi_\mathrm{d} = [-\pi, \pi]$) is performed according to Eq. (3.26). The output is depicted in Figure 3.22.

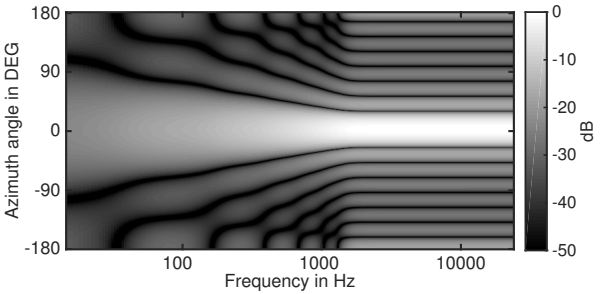


Figure 3.22 Normalized magnitude response $|D^{\mathrm{L}}(\theta_\mathrm{d},\phi_\mathrm{d},\omega)/(N+1)^2|$ of the PWD (regular beampattern) in the azimuth plane ($\theta_\mathrm{d} = \pi/2$, $\phi_\mathrm{d} = [-\pi, \pi]$) versus the temporal frequency.

Below approximately 2 kHz, the PWD looses the CD properties and the beam starts to widen due to missing contributions from higher modal orders, compare Figure 3.21

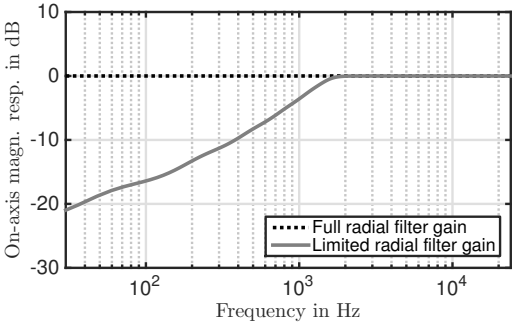


Figure 3.23 Normalized magnitude response $|D^{\mathrm{L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)/(N+1)^2|$ of the PWD for $(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}) = (\theta_{\mathrm{w}}, \phi_{\mathrm{w}})$ with limited radial filter gain. For comparison, the flat response for full radial filter gain is added.

(left plot). Hence, the spatial selectivity of the PWD is sucessively decreased at low frequencies.

Apart from the widened beam, a certain decrease of the on-axis gain can be observed. This is again due to the missing signal contributions from higher modal orders. The on-axis magnitude response of the PWD versus the temporal frequency is depicted in Figure 3.23. The on-axis gain loss can be compensated by amplifying the $n = 0$ omni-directional mode as proposed in (Bernschütz et al., 2011b). No compensation is necessary in this present context as will be discussed in Section 3.6.7.

3.6.5 Non-critical Radial Filters

The question of a reasonable choice for the amplification limit $\hat{a}$ arises. Generally, $\hat{a}$ could be individually adapted according to the signal-to-noise ratio (SNR). However, this approach is unsatisfactory since it does not provide a universally valid solution. The most general choice is lowering the limit to $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$. Here, the radial filters are not allowed to perform any amplification at all. The filters establish the necessary phase relations and perform attenuation only. This way, the typically critical and unstable low frequency range is not critical anymore. Specific restrictions or demands concerning the overall SNR for array processing vanish.

The radial filter magnitudes can be subdivided into two different sections at their minimum around $(\omega/c)\, r_0 = N$, see Figures 3.16 to 3.18. For $(\omega/c)\, r_0 \leq N$ the amplification individually depends on the specific order n and quickly rises to excessive levels. This range is particularly critical. For $(\omega/c)\, r_0 > N$, a common and comparably moderate increase of amplification can be observed that is roughly comparable to

a high-shelf filter. The amplification in this range is less critical for realistic array configurations such as the example configuration depicted in Figure 3.21.

Direct application of the limit of $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ in Eq. (3.25) leads to attenuation in both sections. Attenuation in the upper section $(\omega/c)\, r_0 > N$ should not be applied, since attenuating the amplification in this range would entail a global reduction of high frequencies. Hence, a specific radial filter set is designed, where the soft-knee limiting from Eq. (3.25) is applied for $(\omega/c)\, r_0 \leq N$ only. The upper section $(\omega/c)\, r_0 > N$ is left untouched. The respective filter set is denoted as $\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$.

The magnitude response $|\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ for the example configuration is depicted in Figure 3.24. The scanning PWD according to Figure 3.22 using $\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$ is illustrated in Figure 3.25. The on-axis magnitude is shown in Figure 3.26.

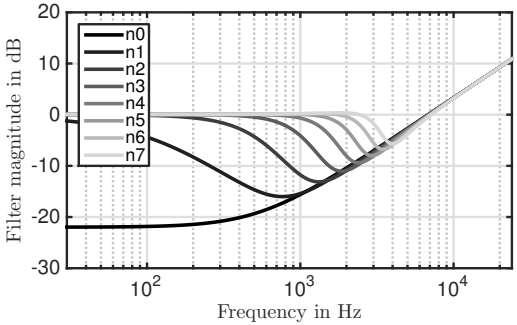


Figure 3.24 Magnitude response $|\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$.

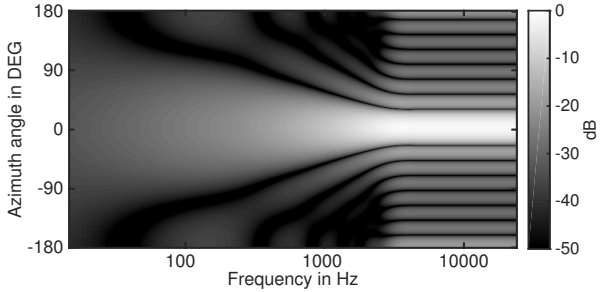


Figure 3.25 Normalized magnitude response $|D^{\mathrm{L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)/(N+1)^2|$ of the PWD (regular beampattern) in the azimuth plane ($\theta_{\mathrm{d}} = \pi/2$, $\phi_{\mathrm{d}} = [-\pi, \pi]$) versus the temporal frequency using $\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$.

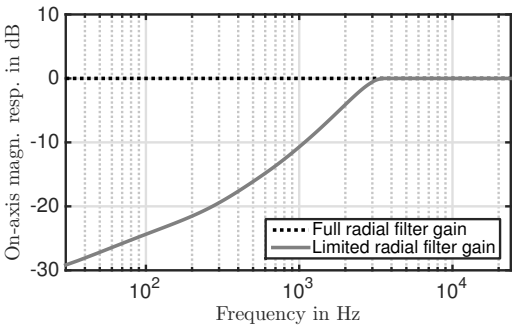


Figure 3.26 Normalized magnitude response $|D^{\mathrm{L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)/(N+1)^2|$ of the PWD for $(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}) = (\theta_{\mathrm{w}}, \phi_{\mathrm{w}})$ with limited radial filter gain using $\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$. For comparison, the flat response for full radial filter gain is added.

3.6.6 Effective Operational Bandwidth (EOB)

At this point we introduce a new measure called effective operational bandwidth (EOB). The EOB refers to the temporal frequency range where the array under certain technical, physical, or numerical constraints still maintains a an ideal response with a certain error margin of e.g. $\pm 3\,\mathrm{dB}$ averaged over all incidence directions. Ideal response refers to the response from an array under idealized conditions.

From Figure 3.25 and Figure 3.26, we can deduce that the EOB referred to $f_{\mathrm{s}}/2 = 24\,\mathrm{kHz}$ for an array of radius $r_0 = 10\,\mathrm{cm}$ and decomposition order $N = 7$ is reduced to around 3.3 oct due to the application of non-critical radial filters.

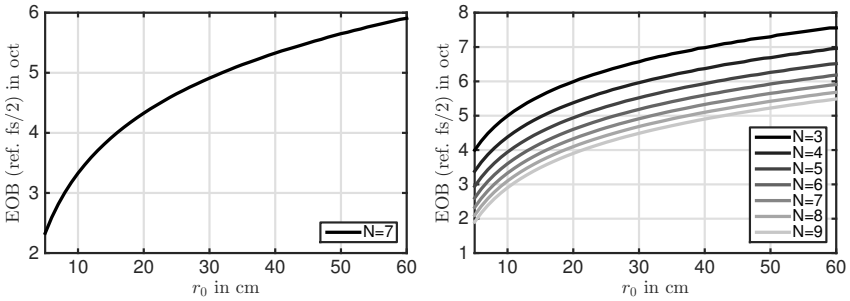


Figure 3.27 Effective operational bandwidth (EOB) referred to $f_{\mathrm{s}}/2$ at a temporal sampling rate of $f_{\mathrm{s}} = 48\,\mathrm{kHz}$ for different orders versus the array radius using non-critical radial filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$.

The EOB using non-critical radial filters ($\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$) depends on the array radius and the decomposition order. Figure 3.27 shows the EOB versus the radius r_0 for $N = 7$ and for $N = [3, 9]$, respectively. The EOB is referred to $f_\mathrm{s}/2 = 24\,\mathrm{kHz}$, since we permit full radial filter amplification in the upper range defined by $(\omega/c)\, r_0 > N$.

A different representation of the same aspect is given in Figure 3.28. The curves denote the absolute lower frequency limit where the array maintains an ideal PWD response. Below this limit the array successively decreases in order and looses spatial resolution. We obtain a drawdown in the magnitude response of the PWD signal.

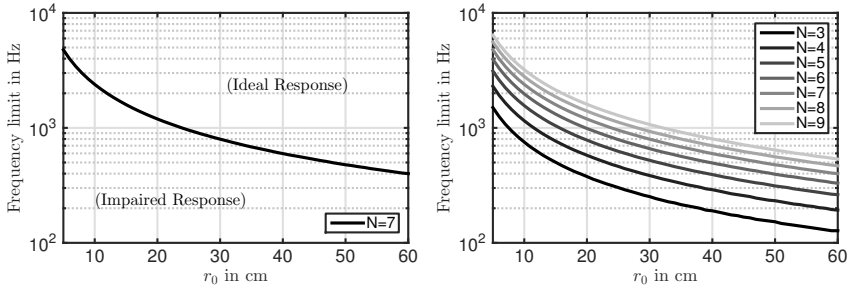


Figure 3.28 Lower frequency limit for an ideal array response versus the array radius using non-critical radial filters $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$.

Finally, we want to know how much bandwidth we gain if we raise the radial filter limit, i.e. $\hat{a}_{\mathrm{dB}} > 0\,\mathrm{dB}$. Figure 3.29 shows the bandwidth gain versus the radial filter limit referred to the non-critical radial filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$.

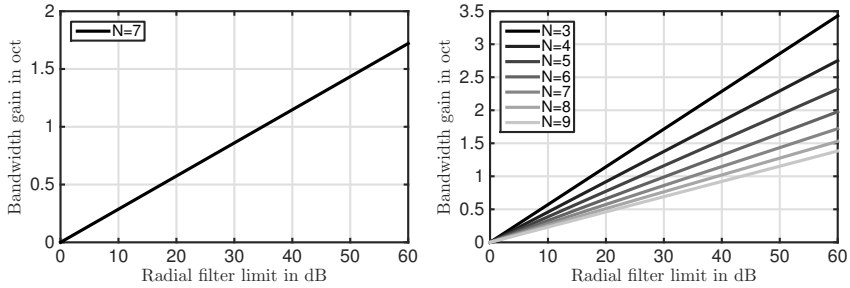


Figure 3.29 Bandwidth gain versus the radial filter amplification limit.

Note that the bandwidth gain depends on the decomposition order and that it increases only moderately when raising the amplification limit. We already need e.g. +35 dB of

additional radial filter gain compared to the non-critical filters in order to achieve a bandwidth gain of 1 oct at a decomposition order of $N = 7$.

3.6.7 Composite Signal

The on-axis magnitude response of the PWD signal shows successive attenuation of low frequencies due to limiting the radial filters, cf. Figure 3.23 and Figure 3.26. In order to achieve a flat on-axis magnitude response of the PWD, the output signal needs to be equalized e.g. by applying an inverted filter that compensates for the attenuation. A different approach is presented in (Bernschütz et al., 2011b). The attenuation of low frequencies is compensated using the $n = 0$ omni-directional signal that provides an excellent SNR in the respective frequency range. However, equalizing the on-axis response has other effects on the diffuse-field response of the array. The array tends to overemphasize diffuse-field components at low frequencies. Defining a suitable equalization filter is difficult, since it depends on the specific context. This is not satisfactory, since we are looking for a general solution.

Fortunately, the apparent problem does not require a solution, when we consider the composite operation of the array. While a single PWD signal shows impairments in the magnitude response, the magnitude of the composite signal $C(\omega)$ (cf. Section 3.8) is not affected by attenuation or impairment due to the radial filter limiting. $C(\omega)$ for the ideal PWD is calculated using

$$C(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}}\, D(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega), \tag{3.27}$$

i.e. by summing several adequately distributed PWD signals on the sphere S. Analogous, we define $C^{\mathrm{L}}(\omega)$ with limited radial filters,

$$C^{\mathrm{L}}(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}}\, D^{\mathrm{L}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega). \tag{3.28}$$

$D^{\mathrm{L}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega)$ is defined in Eq. (3.26). Since the decomposition order N does not influence the magnitude response of $C(\omega)$, cf. Section 3.8, the implicit successive decrease of the modal order does not either. The M_{cg} order-limited PWD responses at $(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}})$ still complement each other, as long as the composite grid fulfills $N_{\mathrm{cg}} \geq N$. We finally find the important relation $C^{\mathrm{L}}(\omega) = C(\omega)$.

We return to the example configuration defined in Section 3.6.2 and the non-critical radial filters from Section 3.6.5. A plane wave $\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_{\mathrm{w}},\phi_{\mathrm{w}})}(r_0, \omega)$ is generated with incidence direction $(\theta_{\mathrm{w}} = \pi/2, \phi_{\mathrm{w}} = 0°)$. Note that the wave incidence direction can

be chosen arbitrarily. A Lebedev composite grid with $M_{\mathrm{cg}} = 86$ nodes yielding $N_{\mathrm{cg}} = 7$ is applied.

The magnitudes of both the composite signal $C^{\mathrm{L}}(\omega)$ and the contributing ($w_{g_{\mathrm{cg}}}$-weighted) PWD signals $D^{\mathrm{L}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega)\, w_{g_{\mathrm{cg}}}$ for this configuration are plotted in Figure 3.30. The partial dropouts in the PWD signals not pointing towards the wave incidence direction are due to the successive order reduction at lower frequencies. The signals cross zero, which does not occur for a PWD with CD. The phenomenon is obvious from Figure 3.22 and Figure 3.25.

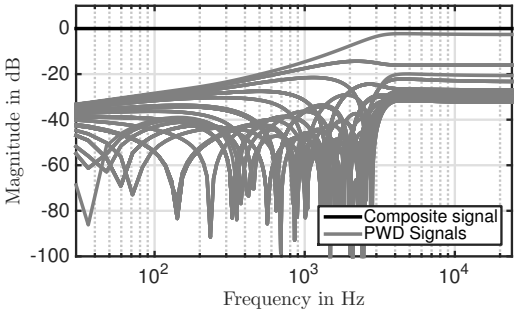


Figure 3.30 Magnitude responses of the composite signal $C^{\mathrm{L}}(\omega)$ and the 86 contributing PWD signals $D^{\mathrm{L}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega)\, w_{g_{\mathrm{cg}}}$.

The composite signal, which is the complex sum of the 86 single PWD signals, shows a perfectly flat magnitude response. Thus, limiting the radial filter amplification of the single PWD signals indeed does not impair the composite signal. This beneficial property of the spherical harmonic de- and recomposition is of fundamental importance for binaural auralization, since we do not need to apply equalization to the low frequency range in order to compensate for the gain loss due to the limiting of the radial filters. The binaural signal is generated using an identical approach but weighting the single PWD signals with HRTFs, cf. Eq. (3.10) or Eq. (3.11).

3.6.8 Binaural Processing

Limiting the radial filter amplification reduces the effective order of the PWD at low frequencies, cf. Figure 3.25. Figure 3.6 shows the modal intensity distribution of the HRTF set. Looking at the lower temporal frequency range $f < 500\,\mathrm{Hz}$, it becomes quickly apparent that the HRTF set does only cover lower orders. The information at low frequencies is largely coded in low orders $n = 0$ up to $n = 2$. For about $f > 1\,\mathrm{kHz}$ higher modal orders start to appear sucessively. It becomes obvious that

at low frequencies there is no need for fully resolving higher modal orders for binaural auralization. As a consequence, the implicit order-reduction due to limiting the radial filters is acceptable under certain conditions.

The critical frequency of the radial filters is around $(\omega/c)\, r_0 \approx N$ and depends on the measurement radius r_0. Non-critical radial filters $|\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ for different measurement radii r_0 are depicted in Figure 3.31.

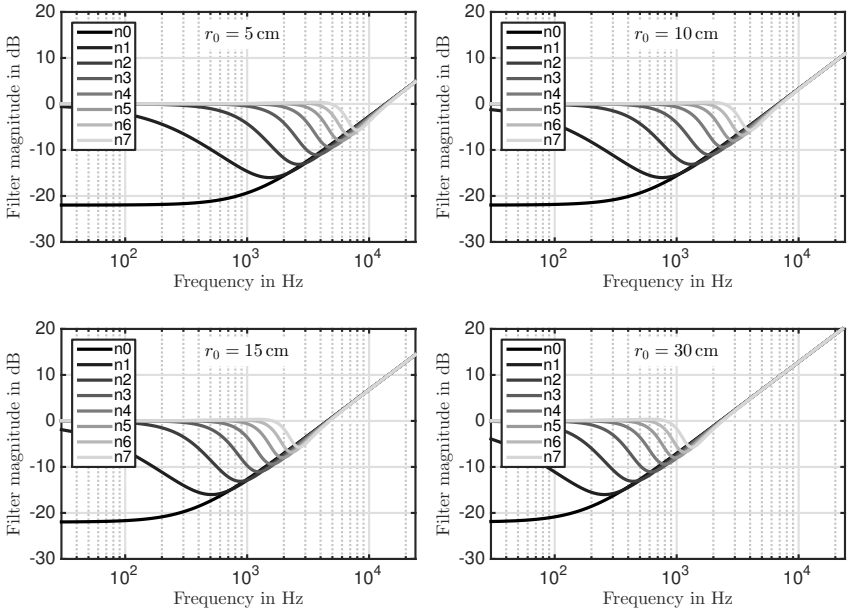


Figure 3.31 Magnitude of the non-critical radial filters $|\tilde{d}_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)|$ for different measurement radii $r_0 = \{5, 10, 15, 30\}\,\mathrm{cm}$.

For $(\omega/c)\, r_0 > N$ full modal resolution N is achievable. For $(\omega/c)\, r_0 < N$ the amplification limiting successively starts to reduce the effective order of the array. For a certain range below $(\omega/c)\, r_0 < N$ the radial filters still maintain their original magnitude until they run into the soft-knee limiter. The limit for fully resolving a mode n is located at the 0 dB-crossing for the original filters or at the knee for the non-critical filters.

If we take the configuration with $r_0 = 15\,\mathrm{cm}$ from Figure 3.31, the critical frequency is located at 2.7 kHz. Mode $n = 7$ is resolved for about $f > 2\,\mathrm{kHz}$, $n = 6$ for approximately $f > 1.5\,\mathrm{kHz}$ or $n = 5$ for about $f > 1\,\mathrm{kHz}$ and so forth.

These values need to be matched with the modal intensity distribution of the HRTF set depicted in Figure 3.6 in order to assess a suitable array radius. An array with around $r_0 \approx 15\,\text{cm}$ turns out to be an optimal choice for non-critical radial filters, since the individual cutoff frequencies match well with the occurrence of respective modal contributions in the HRTF set.

Choosing a radius $r_0 \ll 15\,\text{cm}$ with non-critical radial filters is not feasible, since the decrease of modal resolution due to the amplification limit does not match the intensity distribution of the HRTF set and yields further modal reduction of the HRTFs. A radius of approximately $r_0 \approx 10\,\text{cm}$ is still acceptable, since the decrease in resolution matches comparably well with the HRTFs. By contrast, choosing a radius of e.g. $r_0 = 5\,\text{cm}$ while using non-critical filters is far out of the reasonable range and does not fit for binaural auralization. For a small array radius, the radial filter limits need to be raised until the cutoff frequencies of the mode amplification match the occurrence of the respective modal contributions in the HRTF set. The question of a feasible amplification limit is discussed in the next section, since we first need to understand the propagation of uncorrelated noise through the system.

A radius of $r_0 > 15\,\text{cm}$ does not have a big advantage for binaural auralization, since the resulting improved modal resolution at lower frequencies is not translated by the HRTFs. To the contrary, for bigger radii the required amplification of higher frequencies $(\omega/c)\, r_0 > N$ increases. As a consequence, more equipment noise is amplified in the high frequency range. This can be observed in Figure 3.31 for the $r_0 = 30\,\text{cm}$ radial filter set.

3.7 Noise

For further understanding and optimization, a dedicated analysis of noise propagation through the system from the sensor inputs to the binaural outputs is crucial. Noise refers to uncorrelated white noise that e.g. emanates from the equipment such as microphones, microphone amplifiers, ADCs, and so forth.

We will further see that besides being able to predict the noise response of the system, the dedicated analysis of uncorrelated noise yields a much deeper understanding of the entire approach, especially concerning the interaction between HRTFs and array.

3.7.1 White Noise Gain (WNG)

A common measure for quantifying noise propagation from the input to the output of the array is the white noise gain (WNG), compare e.g. (Rafaely, 2005), (Li and Duraiswami, 2007), (Elko and Meyer, 2009), or (Rettberg and Spors, 2013). The

white noise gain describes the increase of SNR at the system's output referred to its input. Since we treat a linear system here, i.e. the system output is proportional to the system input, the signal can be removed from the observation and the WNG is equivalent to the attenuation of noise from the input to the output. Analogous, WNG^{-1} describes the amplification of noise, which is more intuitive. According to this definition, $\mathrm{WNG}^{-1} > 0\,\mathrm{dB}$ indicates amplification of noise and $\mathrm{WNG}^{-1} < 0\,\mathrm{dB}$ attenuation of noise, respectively.

In the literature, e.g. (Rafaely, 2005), the analysis is mostly restricted to the WNG referred to the output signal of a PWD (or beamformer). This is reasonable, since PWD or beamforming are the predominant applications for arrays. Rettberg and Spors (2013) analyze the impact of noise in spherical beamforming on binaural auralization but again restrict the analysis of the WNG to a single beamformer.

In the specific context of binaural auralization the analysis of the WNG for a single PWD signal is of very limited conclusiveness. Noise propagation through the system is non-trivial. Even though the closed form approach for binaural auralization presented in this thesis is based on performing PWD, the composite signal resulting from the overlay of several PWD signals on the sphere shows substantially different properties than the single contributing PWD signals. Furthermore, the HRTF set acts as a complex modal filter in the middle of both and has impact on the WNG. Thus, analyzing the properties of a single PWD or beamformer is not sufficient for assessing the overall system properties. In order to gain a deeper insight, we analyze a PWD signal, the composite signal, and the binaural output signal for varying constellations in the following.

We are mainly interested in the expected WNG of the binaural output signal referred to the array inputs and how array size, radial filter limit, number of sensors, or decomposition order influence the overall WNG.

3.7.2 Analytic Description

Be $\gamma(t)$ Gaussian white noise with variance $\sigma^2 = 1$. Its time-frequency Fourier transform is $\Gamma(\omega) = \mathcal{F}_\mathrm{t}\{\gamma(t)\}$, where $\Gamma(\omega)$ has a constant power spectral density. Gaussian white noise is a reasonable assumption for equipment noise.

Since every input signal path of the array produces an independent realization of $\gamma(t)$, we extent the expression to $\gamma_g(t)$, where g denotes a single realization corresponding to a specific grid node g of the spatial sampling grid. The realizations $\gamma_g(t)$ and $\gamma_{g+1}(t)$ are uncorrelated.

The noise realizations at nodes g are transformed to the spherical wave spectrum using the discrete spatial Fourier transform according to Eq. (3.3).

$$\mathring{\Gamma}_{nm}(\omega) = \sum_{g_{\mathrm{sg}}=1}^{M_{\mathrm{sg}}} w_{g_{\mathrm{sg}}}\, \Gamma_g(\omega)\, Y_n^m(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})^*, \tag{3.29}$$

where $\Gamma_g(\omega)$ describes a noise realization at node g in the space-frequency domain and $\mathring{\Gamma}_{nm}(r_0, \omega)$ the spatial Fourier coefficients in the spherical wave spectrum domain. $g \in [1, M_{\mathrm{sg}}]$ indicates the node and $w_{g_{\mathrm{sg}}}$ are is respective grid weight of the spatial sampling grid.

To simplify matters, we restrict the analysis to rigid sphere array configurations and their related radial filters $d_n^{\mathrm{RS}}(\frac{\omega}{c}\, r_0)$, cf. Section 3.6.

The output of a single PWD signal is calculated using

$$D^{\mathrm{N}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) = 4\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0)\, \mathring{\Gamma}^u_{nm}(\omega)\, Y_n^m(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}), \tag{3.30}$$

The composite signal is generated by weighted summation of multiple PWD signals defined by the nodes g_c of a suitable composite grid, according to Eq. (3.8) and Section 3.8. The composite signal $C^{\mathrm{N}}(\omega)$ is obtained using

$$C^{\mathrm{N}}(\omega) = \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}}\, D^{\mathrm{N}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega). \tag{3.31}$$

For performing noise analysis with respect to the binaural output, we only consider a single ear (left ear). Assuming symmetrical anatomy of the head we expect identical results for both ears. The binaural output is generated analogous to Eq. (3.11) using

$$Y^{\mathrm{N}}(\omega) = \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}} \Bigg[\sum_{g'=1}^{M'} \sum_{n'=0}^{N'} \sum_{m'=-n'}^{n'} w'_{g'}\, H^{\mathrm{l}}(\theta_{g'}, \phi_{g'}, \omega)\, Y_n^m(\theta_{g'}, \phi_{g'})^* \times Y_n^m(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}) \Bigg] D^{\mathrm{N}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega). \tag{3.32}$$

The reference input noise for the WNG calculation is defined by

$$\overline{\gamma}(t) = \frac{1}{\sqrt{M_{\mathrm{sg}}}} \sum_{g_{\mathrm{sg}}=1}^{M_{\mathrm{sg}}} \gamma_g(t), \tag{3.33}$$

with its corresponding frequency domain representation $\overline{\Gamma}(\omega)$.

For assessing the WNG of the PWD we can observe arbitrary decomposition directions, e.g. ($\theta_\mathrm{d} = pi/2, \phi_\mathrm{d} = 0$). We are interested in the specific contribution of a single PWD signal related to the composite signal or the binaural output signal. Since the contributing PWD signals are weighted with individual node weights that are defined by the composite grid, every single PWD signal is individually attenuated by its assigned weight. In order to appropriately account for the reduced sensitivity of a single PWD contribution, we weight the PWD signal with a mean composite grid weight $\overline{w}_\mathrm{cg}$, which is equivalent to the reciprocal of the number of nodes for all quadratures, i.e. $\overline{w}_\mathrm{cg} = 1/M_\mathrm{sg}$. This yields a representative relation of the PWD signal compared to the composite signal or the binaural signal. Accounting for the average weight $\overline{w}_\mathrm{cg}$ and the freely defined decomposition angle ($\theta_\mathrm{d} = \pi/2, \phi_\mathrm{d} = 0$) we obtain a slightly modified version of Eq. (3.30),

$$\check{D}^\mathrm{N}(\omega) = 4\pi\,\overline{w}_\mathrm{cg} \sum_{n=0}^{N} \sum_{m=-n}^{n} d_n^\mathrm{RS}(\tfrac{\omega}{c}\, r_0)\, \mathring{\Gamma}_{nm}^{u}(\omega)\, Y_n^m(\pi/2, 0), \tag{3.34}$$

where $\check{D}^\mathrm{N}(\omega)$ denotes a representative weighted PWD output signal. Finally, the WNG (in dB) of the weighted PWD signal is obtained using

$$\mathrm{WNG}_{\check{\mathrm{D}}}(\omega) = 10\log_{10}\frac{|\overline{\Gamma}(\omega)|^2}{|\check{D}^\mathrm{N}(\omega)|^2}. \tag{3.35}$$

Analogous, the calculation of the WNGs for the composite signal and the binaural signal yields

$$\mathrm{WNG}_\mathrm{C}(\omega) = 10\log_{10}\frac{|\overline{\Gamma}(\omega)|^2}{|C^\mathrm{N}(\omega)|^2}, \tag{3.36}$$

and

$$\mathrm{WNG}_\mathrm{Y}(\omega) = 10\log_{10}\frac{|\overline{\Gamma}(\omega)|^2}{|Y^\mathrm{N}(\omega)|^2}, \tag{3.37}$$

respectively.

3.7.3 WNG Analysis

Simulations are performed in order to determine the different noise transmission functions. In practice, a single realization of white noise with finite signal length does usually not show constant power spectral density as previously assumed for an infinite signal in theory. In order to obtain representative simulation results, we sequentially feed several independent realizations into the processing chain and average the magnitudes at the respective outputs, as well as the magnitudes of the reference noise. The

averaged power spectral density becomes more constant as we increase the repetitions. For the subsequent simulations, 200 uncorrelated realizations with 2^{14} random samples are averaged yielding an improved constancy of the power spectral density with an acceptable amount of residual ripple. In a final step, the WNGs are smoothed with $1/6$ oct.

We start the WNG analysis with our example configuration using a rigid sphere array with $N = 7$, $r_0 = 10\,\mathrm{cm}$, $M_{\mathrm{sg}} = 86$, and non-critical radial filters (amplification limit $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$). We use Lebedev quadratures for spatial sampling and order-matched Gauss quadratures as composite grids.

Since we aim to analyze the impact of changing dedicated parameters, the given configuration is used as arbitrarily chosen anchor or reference configuration in the following. Figure 3.32 shows the WNG^{-1} of the weighted PWD signal $\mathrm{WNG}_{\tilde{\mathrm{D}}}^{-1}(\omega)$, the composite signal $\mathrm{WNG}_{\mathrm{C}}^{-1}(\omega)$ and the binaural signal $\mathrm{WNG}_{\mathrm{C}}^{-1}(\omega)$ for the given setup.

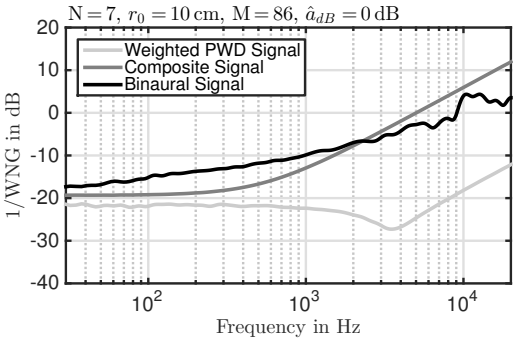


Figure 3.32 WNG^{-1} of the weighted PWD signal $\mathrm{WNG}_{\tilde{\mathrm{D}}}^{-1}(\omega)$, the composite signal $\mathrm{WNG}_{\mathrm{C}}^{-1}(\omega)$ and the binaural signal $\mathrm{WNG}_{\mathrm{Y}}^{-1}(\omega)$ for the reference configuration.

From Figure 3.32 we can derive some basic observations. First, the WNG of the PWD signal closely follows the shape of the radial filter of the highest order involved, which is $n = 7$ in the present case. Hence, the WNG of the single PWD is essentially determined by the highest order of decomposition. Lower orders indeed contribute as well but are less influential due the fact that they contribute with lower amplitudes. The global attenuation of around $-20\,\mathrm{dB}$ is due to the applied mean node weight $\overline{w}_{\mathrm{cg}} = 1/M_{\mathrm{sg}}$ that yields an attenuation of $10\log 10(1/M_{\mathrm{sg}})\,\mathrm{dB}$. The run and absolute magnitude of $\mathrm{WNG}_{\tilde{\mathrm{D}}}^{-1}(\omega)$ conforms with observations described in literature e.g. (Rafaely, 2005). Note that for direct comparison to literature, the specific weighting $\overline{w}_{\mathrm{cg}} = 1/M_{\mathrm{sg}}$ and the introduced amplitude limiting need to be removed.

Focusing on the WNG of the composite signal, we observe an entirely different behavior. The WNG of the composite signal follows the run of the radial filter for $n = 0$. Keeping in mind that the composite signal represents the composed omni-directional signal at the origin, and, that $n = 0$ addresses the omni-directional mode of the spherical harmonics, this result appears to be plausible but is certainly not trivial.

In order to understand the differences between the WNG of the binaural signal $\mathrm{WNG_Y}$ and the WNG of the composite signal $\mathrm{WNG_C}$ we modify some of the simulations parameters.

3.7.3.1 Decomposition Order and Number of Spatial Sampling Nodes

In this section we study the question whether the decomposition order has an influence on the WNG. For this purpose, we vary the decomposition order. At the same time the number of sampling nodes is adapted in order to fulfill $N_{\mathrm{sg}} = N$. Figure 3.33 shows the results for orders $N = \{5, 7, 9, 11\}$ that demand a minimum of $M_{\mathrm{sg}} = \{50, 86, 146, 194\}$ spatial sampling nodes of the Lebedev sampling quadrature.

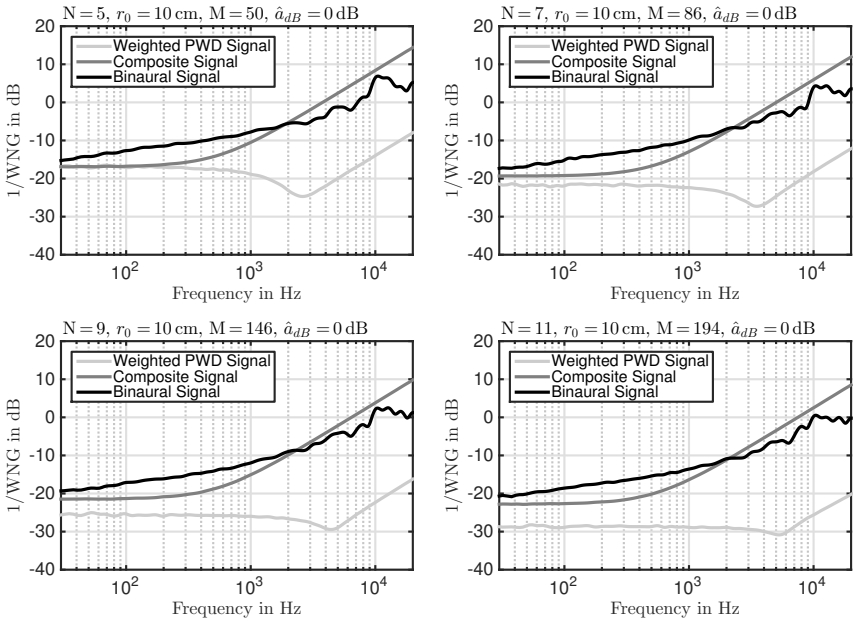


Figure 3.33 $\mathrm{WNG}^{-1}_{\mathrm{D,C,Y}}$ for different decomposition orders $N = \{5, 7, 9, 11\}$ using spatial sampling grids that fulfill $N_{\mathrm{sg}} = N$ with $M_{\mathrm{sg}} = \{50, 86, 146, 194\}$ nodes.

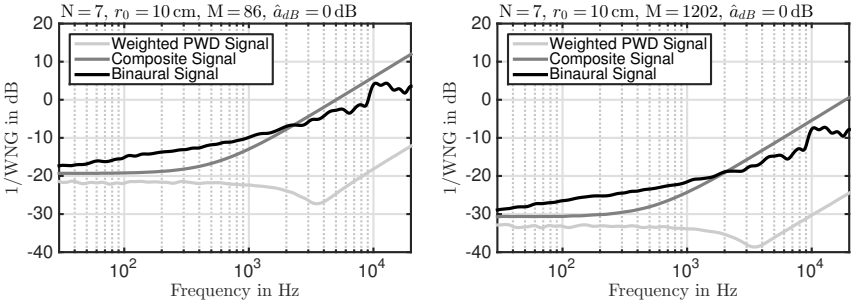


Figure 3.34 $\mathrm{WNG}^{-1}_{\breve{\mathrm{D}},\mathrm{C},\mathrm{Y}}$ for decomposition orders $N = 7$ using different spatial sampling grids with $M_{\mathrm{sg}} = \{86, 1202\}$ nodes (Lebedev). Both grid fulfill $N_{\mathrm{sg}} \geq N$.

From the figure we can derive that $\mathrm{WNG}^{-1}_{\breve{\mathrm{D}}}(\omega)$ is determined by the radial filter of the highest involved mode $n = N$.

In contrast to the PWD $\mathrm{WNG}^{-1}_{\breve{\mathrm{D}}}(\omega)$, the composite signal $\mathrm{WNG}_{\mathrm{C}}(\omega)$ is independent of the decomposition order. The same holds true for the binaural signal. There are some minor order-dependent fluctuations in the high frequency range that can be ascribed to the order-reduction artifacts of RHRTFs discussed in Section 3.5.2. Otherwise we see that $\mathrm{WNG}_{\mathrm{Y}}(\omega)$ is independent of the decomposition order N.

We also find that the overall magnitudes scale with $1/\sqrt{M_{\mathrm{sg}}}$, i.e. they depend on the number of sensors. In order to assure the latter assumption, we modify our reference design. We keep the decomposition order $N = 7$ constant but dramatically increase the number of sensors to $M_{\mathrm{sg}} = 1202$. The result is depicted in Figure 3.34.

Now, everything scales perfectly with $1/\sqrt{M_{\mathrm{sg}}}$. As a consequence, we can be sure the observed scaling in Figure 3.33 is solely ascribed to the number of sensors and there is no influence of the decomposition order on $\mathrm{WNG}_{\mathrm{C}}(\omega)$ and $\mathrm{WNG}_{\mathrm{Y}}(\omega)$, if we neglect the observed minor fluctuations at high frequencies that are due to the RHRTF order-reduction artifacts observed in Figure 3.33.

3.7.3.2 Array Radius

In order to evaluate the influence of the array radius on the WNG and to further understand the relation between WNG of the composite signal and WNG of the binaural signal, we vary the radius of the reference design, $r_0 = \{5, 10, 20, 40\}$ cm. The respective curves are depicted in Figure 3.35. Since we see a multitude of changes emerging at once, we discuss the observations for the PWD signal $\mathrm{WNG}_{\breve{\mathrm{D}}}$, the composite signal $\mathrm{WNG}_{\mathrm{C}}$, and the binaural signal $\mathrm{WNG}_{\mathrm{Y}}$ step-by-step.

3.7.3.2.1 WNG of the PWD The behavior of $\mathrm{WNG}_{\breve{\mathrm{D}}}^{-1}$ does not reveal new information. The curves remain akin to the respective radial filter of the highest modal order, i.e. $n = 7$, consistent with earlier observations. The typical knee that is located around $\omega = N\,c/r_0$ moves along the frequency axis in proportion to the measurement radius r_0. This is just analogous to the radial filters.

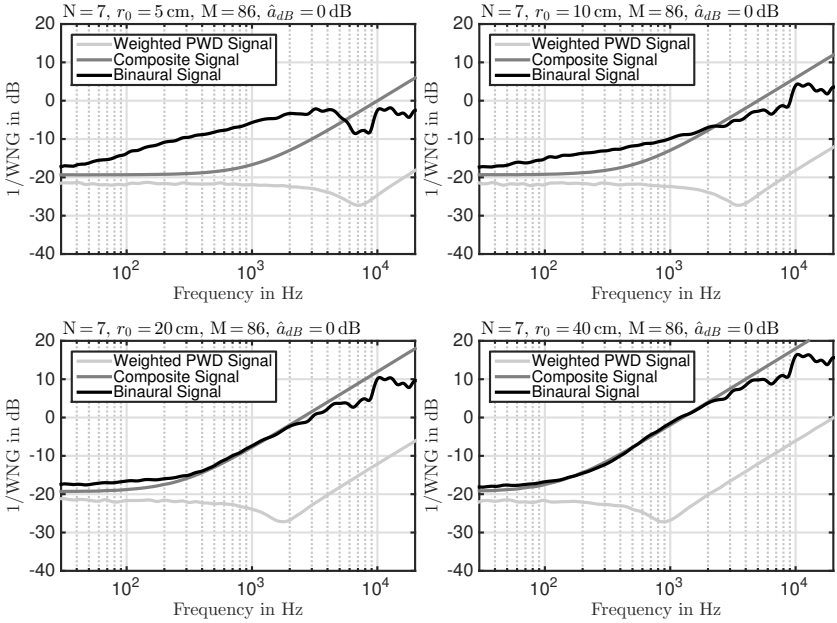


Figure 3.35 $\mathrm{WNG}_{\breve{\mathrm{D}},\mathrm{C},\mathrm{Y}}^{-1}$ for different measurement radii $r_0 = \{5, 10, 20, 40\}$ cm.

3.7.3.2.2 WNG of the Composite Signal Likewise, the WNG of the composite signal moves along the frequency axis in proportion to the measurement radius r_0. We further analyze the particular structure of $\mathrm{WNG}_{\mathrm{C}}^{-1}$. The $\mathrm{WNG}_{\mathrm{C}}^{-1}$ can be subdivided into two different sections. There is a constant section for $\omega < (c/r_0)$ and an ascending section for $\omega > (c/r_0)$. If we account for the damping factor of $1/\sqrt{M_{\mathrm{sg}}}$, we come to the important finding that a spherical array operating in the underlying composite mode provides a natural WNG according to

$$\mathrm{WNG}_{\mathrm{C}}(\omega) = 10\,\log_{10}\left[\frac{M_{\mathrm{sg}}}{1 + \left(\frac{\omega\,r}{c}\right)^2}\right]\mathrm{dB}, \tag{3.38}$$

which in terms of noise amplification, i.e. $\mathrm{WNG}_{\mathrm{C}}^{-1}$, yields

$$\mathrm{WNG}_{\mathrm{C}}^{-1}(\omega) = 10\,\log_{10}\left[\frac{1+\left(\frac{\omega\,r}{c}\right)^2}{M_{\mathrm{sg}}}\right]\mathrm{dB}. \tag{3.39}$$

The noise amplification $\mathrm{WNG}_{\mathrm{C}}^{-1}$ results from an overlay of a constant term (0 dB) and an ascending term (6 dB/oct). Both terms are scaled by $1/\sqrt{M_{\mathrm{sg}}}$. Thus, for every doubling of the number of transducers, the resulting power spectral density at the composite array output decreases by 3 dB. The noise amplification as a function of the ratio $(\omega/c)\,r_0$ is depicted in Figure 3.36.

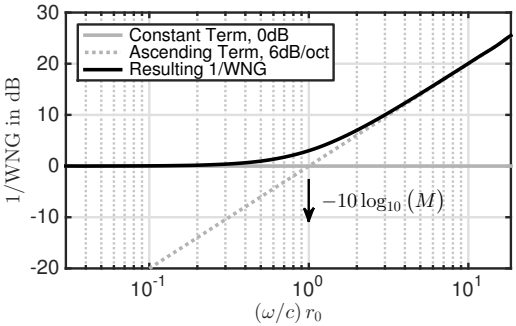


Figure 3.36 Natural WNG^{-1} of a spherical array in composite mode referred to a single transducer ($M_{\mathrm{sg}} = 1$). The entire curve is shifted down by $10\,\log_{10}(\sqrt{M_{\mathrm{sg}}})$ dB in order to account for the actual number of transducers in the array.

The WNG of the composite signal is independent of the radial filter amplification limit, which conforms with the findings in Section 3.6.7.

3.7.3.2.3 WNG of the Binaural Signal The WNG of the binaural signal is more difficult to understand. In Figure 3.35 we observe that the WNG of the binaural signal converges to the WNG of the composite signal for larger array radii. We also see some irregular drawdown at higher frequencies, $f > 2$ kHz. This needs to be discussed first.

We determine the diffuse-field response $H^{1,\mathrm{DF}}(\omega)$ of the spherical HRTF set (left ear) by integrating the magnitudes of the HRTFs over the entire sphere,

$$H^{1,\mathrm{DF}}(\omega) = \frac{1}{4\pi}\iint_S |H^{1}(\theta,\phi,\omega)|\,\mathrm{d}\Omega. \tag{3.40}$$

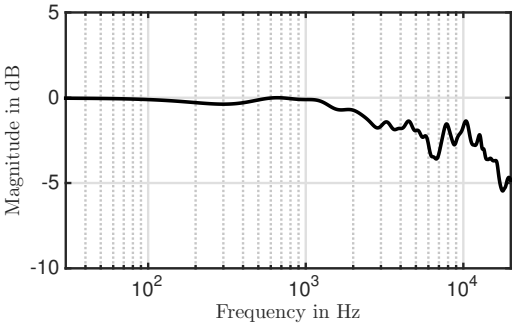


Figure 3.37 Diffuse-field response $H^{\mathrm{l,DF}}(\omega)$ of the HRTF set from Section 4.3.2.

The resulting diffuse-field response $H^{\mathrm{l,DF}}(\omega)$ is depicted in Figure 3.37. From this figure we understand the origin of the irregular drawdown at higher frequencies in $\mathrm{WNG}_{\mathrm{Y}}^{-1}$; it corresponds to the native diffuse-field response of the HRTF set.

We also need to keep in mind the inherent order-reduction of the HRTFs leading to RHRTFs, as discussed in Section 3.5.2. The order-reduction introduces errors that distinguish RHRTFs from HRTFs. These errors not only influence single RHRTFs, but also impact the overall diffuse-field response of the entire RHRTF set.

As a consequence, we need to gather the diffuse-field response of the RHRTFs $H_N^{\mathrm{l,r}}$ rather than the diffuse-field response of the original HRTFs $H^{\mathrm{l,r}}$ yielding

$$H_N^{l,\,\mathrm{DF}}(\omega) = \frac{1}{4\pi} \iint_S |H_N^{\mathrm{l}}(\theta, \phi, \omega)| \, \mathrm{d}\Omega. \tag{3.41}$$

This difference explains some deviations, as well as the minor fluctuations in the upper frequency range that are observed in Figure 3.33, where different decomposition orders are compared.

So far we understand the phenomena that emerge in the high frequency range. In a next step we concentrate on the low frequency range.

From Figure 3.35 we can see that the noise amplification $\mathrm{WNG}_{\mathrm{Y}}^{-1}$ of the binaural signal raises for $\omega = N\,c/r_0$ when decreasing the array radius. In contrast, the WNG of the binaural signal tends to converge to the WNG of the composite signal with increasing array radius.

If we consider that a HRTF set is widely omni-directional at low temporal frequencies, we understand why both, $\mathrm{WNG}_{\mathrm{Y}}(\omega)$ and $\mathrm{WNG}_{\mathrm{C}}(\omega)$, tend to show similar characteristics. The important difference is the spatial offset of the ear from the physical origin, which is discussed in Section 3.5.

While the composite signal represents an omni-directional signal at the physical and mathematical origin, the binaural signal emerges from a position with a spatial offset from the origin that further varies with the wave incidence direction. This spatial offset excites higher spherical harmonic orders $n > 0$ even at low frequencies where the HRTF set could otherwise be assumed fully omni-directional, compare Figure 3.6.

Due to the radial filter limiting higher modes are successively suppressed at low temporal frequencies. Suppressing higher modal orders entails truncation errors in the modal representation of the HRTFs. The truncation error yields increased noise levels in the affected range, since the higher modal contributions do not cancel, which is the case in the composite signal. Increasing the radius, the higher modes at low temporal frequencies undercut the radial filter limit and become available to resolve the HRTFs properly. This decreases the truncation error and yields an increased WNG of the binaural signal, whereby the natural upper limit is generally determined by the WNG of the composite signal.

So far we conclude that using non-critical radial filters ($\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$) and an array size $r_0 \to \infty$, the binaural signal has a $\mathrm{WNG_Y}$ according to

$$\mathrm{WNG_Y}(\omega) = 10\,\log_{10}\left[\frac{M_{\mathrm{sg}}}{1+\left(\frac{\omega\,r}{c}\right)^2}\right] - 20\log_{10}\left[H_N^{l,\,\mathrm{DF}}(\omega)\right]\mathrm{dB}, \tag{3.42}$$

which corresponds to the WNG of the composite signal weighted by the diffuse-field response of the RHRTFs.

In practice, we can neglect the order truncation effects in the diffuse-field response and roughly approximate $\mathrm{WNG_Y}$ for an array that is greater or equal than the height of a human head, i.e. $r_0 \geq r_{h_{\mathrm{head}}}$, by

$$\mathrm{WNG_Y} \approx 10\,\log_{10}\left[\frac{M_{\mathrm{sg}}}{1+\left(\frac{\omega\,r}{c}\right)^2}\right] - 20\log_{10}\left[H^{l,\,\mathrm{DF}}(\omega)\right]\mathrm{dB}. \tag{3.43}$$

3.7.3.3 Radial Filter Amplification Limit

Since we have seen that the radial filter limiting entails truncation errors that yield increased noise levels, we might want to increase the radial filter limit. We set the radial filter limit to $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$ for the reference configuration. The result is shown in Figure 3.38.

The PWD signal $\mathrm{WNG}_{\check{\mathrm{D}}}^{-1}$ responds proportional to the raised radial filter limit and we observe the expected increase of $\mathrm{WNG}_{\check{\mathrm{D}}}^{-1}$ at low frequencies. The composite signal $\mathrm{WNG_C^{-1}}$, by contrast, is actually independent of the radial filter limit.

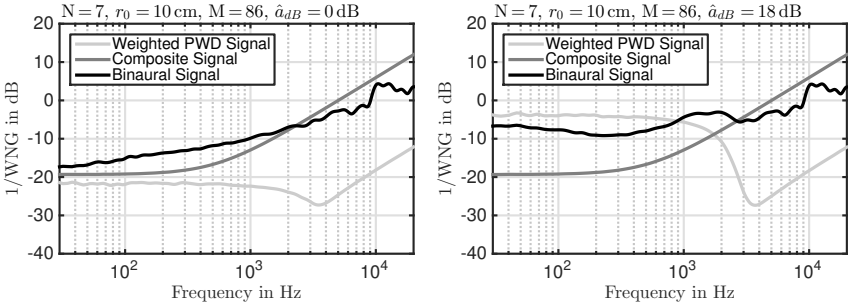


Figure 3.38 $\mathrm{WNG}^{-1}_{\tilde{\mathrm{D}},\mathrm{C},\mathrm{Y}}$ for non-critical radial filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ and radial filters with $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$.

The binaural signal $\mathrm{WNG}^{-1}_{\mathrm{Y}}$ is indeed affected but does not respond proportional to the raised radial filter gain. We observe this ambivalent response, because the truncation error is reduced due to the availability of higher modal orders, while the remaining truncation error residuals leak the increased noise power delivered by the single PWD signals. Unfortunately, the decreased truncation error does not compensate for the leaked noise, which leads to increased noise levels.

In theory, we could omit the limiting of the radial filter amplification to provide full modal resolution and avoid the truncation errors. In this case, the WNG of the binaural signal equals the WNG of the composite signal. In practice, however, opening the radial filters implicates that even smallest errors (sensor differences, sensor positioning errors, etc.) are amplified excessively and the output signal immediately blows up. Since we deal with amplification ratios in the range of several hundred decibels at low frequencies, we even may reach the numerical limit of the variables in the processing software. In order to avoid instability and to obtain a robust array response, radial filter amplification limiting is absolutely indispensable in practice.

We now increase the array radius and use the raised radial filter limit of $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$. The result is depicted in Figure 3.39.

As long as the modal HRTF filters avail of all required modes, the truncation error vanishes and the WNG of the binaural signal becomes equal to the WNG of the composite signal. In this specific range the WNG of the binaural signal turns out to be equally independent of the radial filter gain. However, increased output noise appears in the range where modal truncation is applied. Hence, the WNG of the binaural signal turns out to depend on the radial filter amplification limit, but in a decisively different manner than the WNG of the PWD.

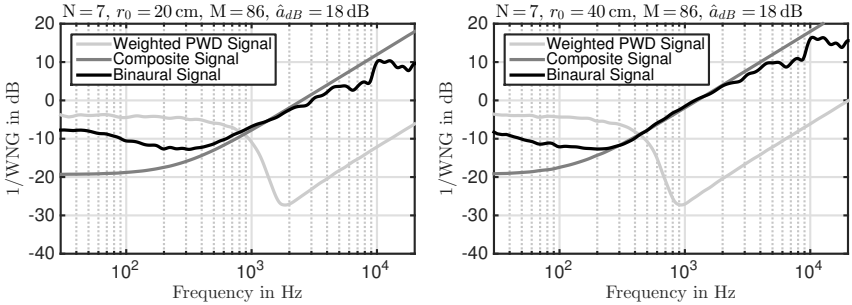


Figure 3.39 $\mathrm{WNG}^{-1}_{\tilde{\mathrm{D}},\mathrm{C},\mathrm{Y}}$ for radii $r_0 = \{20, 40\}$ cm using radial filters with an amplification limit of $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$.

We see that raising the radial filter limit and decreasing the radius yield increased noise amplification in the low frequency range of the binaural signal. As a consequence, the proposed non-critical radial filters turn out to be optimal in terms of the WNG.

There is a certain WNG headroom, up to $10\log_{10}(M_{\mathrm{sg}})$ dB in the low frequency range, that can be exploited for raising the radial filter amplification limit and for increasing the effective operational bandwidth (EOB) of the array, refer to Section 3.6.6, while maintaining a reasonable overall WNG. The increased bandwidth allows for decreasing the array radius while still resolving the HRTFs properly, refer to Section 3.6.8. Decreasing the array radius is beneficial, as it reduces noise in the higher frequency range, where the WNG of the binaural signal is low.

3.7.3.4 Optimization of Array Parameters

Optimizing the array radius for binaural auralization consists of finding a suitable trade-off between noise at higher frequencies (larger radius → more noise) and noise at low frequencies (smaller radius or raised radial filter limit → more noise), while simultaneously preserving an adequate modal resolution of the HRTFs according to their modal intensity distribution in the spherical wave spectrum domain, as discussed in Section 3.6.8. The optimization also depends on the actual number of transducers used in the array.

The optimization task must be individually performed depending on the total number of sensors in the array. In practice, we find a very manageable margin for reasonable array parameters. We strive to use as few sensors as possible, since additional microphones (plus microphone amplifiers and ADCs) increase the costs. Therefore, we tend to use the minimum number of nodes to fulfill the basic requirement $N_{\mathrm{sg}} = N$.

Thus, if we aim for a decomposition order of $N = 7$ for example, we apply $M_{\mathrm{sg}} = 86$ microphones arranged on a Lebedev sampling grid to satisfy $N_{\mathrm{sg}} = N$. For this configuration an array radius of $r_0 = 10\,\mathrm{cm}$ with $M_{\mathrm{sg}} = 86$ transducers ($N_{\mathrm{sg}} = 7$) and an amplification limit of $\hat{a}_{\mathrm{dB}} \approx 20\,\mathrm{dB}$ keeps a relatively good balance between the effective operational array bandwidth and WNG constraints at low and high temporal frequencies. This configuration is similar to the one that is analyzed in Figure 3.38 (right plot).

Increasing the radius yields more noise at high frequencies. Decreasing the radius and raising the amplification limit yields more noise at low frequencies. Furthermore, decreasing the radius while not raising the amplification limit yields modal truncation of the HRTF set in the mid-frequency range (just below $\omega < (c/r_0)$), which impairs the RHRTF response.

In general, we find the optimum trade-off between the parameters when $\mathrm{WNG}_{\mathrm{Y}}$ is close to 0 dB and as constant as possible. We restrict ourselves to as constant as possible instead of constant, since the WNG does not become perfectly constant and the output noise always remains colored. Hence, the design target is to achieve a $\mathrm{WNG}_{\mathrm{Y}}$ of around 0 dB with minimum fluctuation over the temporal frequency. We conclude that if we feed uncorrelated white noise to the array inputs, we do not get white noise at the binaural output, but noise with a certain spectral coloration depending on the array parameters and the diffuse-field response of the HRTF set.

3.7.4 Conclusions

To conclude this section, we summarize some of the major aspects of the WNG:

- Analyzing the WNG of the PWD only, which is the most common approach in literature, turns out to be inconclusive for assessing the WNG properties of an array-based binaural system. The composite approach needs to be considered.
- Neglecting minor order-reduction artifacts of the HRTFs at high temporal frequencies, the WNG of the binaural system is independent of the decomposition order.
- For large array radii (e.g. $r_0 > 15\,\mathrm{cm}$) the WNG of the binaural output tends to follow the WNG of the composite output. Additionally, it is weighted with the diffuse field response of the (R)HRTF set.
- For small array radii (e.g. $r_0 < 15\,\mathrm{cm}$) by contrast, the WNG rapidly decreases for temporal frequencies $\omega < N\,c/r_0$ when compared to the composite signal.

- Radial filter amplification limiting is indispensable for obtaining a robust array response in practice. The presented non-critical radial filters ($\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$) yield highest possible WNG over the entire temporal frequency spectrum.
- The WNG of the composite signal is generally unaffected by the radial filter amplification limiting. However, the WNG of the binaural signal depends on the radial filter amplification limit, but in a different way than the WNG of the PWD.
- We found analytic expressions for predicting the composite WNG, cf. Eq. (3.38), as well as the binaural WNG, cf. Eq.(3.42) and Eq.(3.43).
- The WNG of an array in composite mode turns out to be constant for $\omega < (c/r_0)$ and shows a decrease of 6 dB/oct for $\omega > (c/r_0)$. The overall WNG is scales with the number of sensors in the array.
- Raising the radial filter amplification limit within reasonable boundaries (e.g. up to +20 dB) increases the effective operational bandwidth (EOB) and permits to decrease the array radius slightly.
- Finding optimum array parameters means to find the optimum trade-off between array radius, number of sensors, and radial filter gain limit. The design target is to obtain noise at the output that provides the best possible constancy of the power spectral density.
- The optimum array size for performing array-based binaural auralization turns out to be more or less the size of a human head, i.e. between approximately $r_0 = 10\,\mathrm{cm}$ and $r_0 = 12\,\mathrm{cm}$. Smaller arrays raise the noise at low temporal frequencies and tend to truncate the modal order of the HRTF set in a crucial band. Larger arrays, by contrast, raise the noise at high temporal frequencies. The gained modal resolution at lower temporal frequencies dissipates, since it is not resolved by the HRTFs in a useful way.
- The resulting noise at the binaural output is colored depending on the array size, the radial filter limit, and the diffuse field response of the HRTF set. Even a well-optimized system can be expected to show some residual coloration of noise at the binaural outputs.

3.8 Spatial Aliasing

Spatial aliasing or angular aliasing is a consequence of discrete spatial sampling. Since the structure of the spherical harmonics gets progressively more complex at higher orders, any discrete sampling scheme reaches a certain limit of resolution, which was

previously denoted by N_{sg} in Section 3.2.2. For spherical harmonic orders $N > N_{\mathrm{sg}}$ the modes cannot be uniquely identified.

As long as the function on the sphere is band-limited to $N \leq N_{\mathrm{sg}}$, no spatial aliasing arises, since the information can be uniquely decoded by reading the different modes. Since natural sound fields are not band-limited, spatial aliasing is expected. Higher modes project aliases into lower modes due to identification ambiguities. The isolated information contained in the dedicated modes is mixed up. Spatial aliasing decreases the spatial resolution of the system and can impede reasonable decomposition under certain conditions.

The number and arrangement of sampling nodes for an achievable grid target order N_{sg} is discussed in Section 3.2.2. The radius r_0 of the sphere S_0 that embeds the sampling nodes, as well as the analyzed temporal frequency within the sampled wave field must be taken into account..

Keeping the number and arrangement of sampling nodes constant while changing the radius r_0 influences the effective node density on S_0. Additionally, higher temporal frequencies in the sound field progressively tend to excite higher spherical harmonic modes on S_0. As a consequence, high temporal frequencies are more prone to spatial aliasing than low temporal frequencies. Sampling a natural sound field using discrete sampling positions always brings up a certain spatial aliasing contribution throughout the entire time-frequency spectrum. The contributions are less intense at lower temporal frequencies and progressively more intense for higher temporal frequencies. Thus, there is no dedicated frequency boundary for a spatial aliasing-free operation in the time-frequency domain that would be equivalent to the Nyquist-frequency (Nyquist, 1928), (Shannon, 1949), known from discrete temporal sampling. The true equivalent to the Nyquist-frequency would refer to the maximum modal order, i.e. to N_{sg}, and not to a temporal frequency, since spatial aliasing refers to the space domain / spherical wave spectrum domain and only implicitly affects the time-frequency domain.

However, we can indeed observe a certain temporal frequency, where spatial aliasing contributions start to surge excessively. This specific temporal frequency is denoted as spatial aliasing frequency[3], f_{A}. It can be estimated using (Rafaely, 2005) (Rafaely et al., 2007b)

$$f_{\mathrm{A}} = \frac{N_{\mathrm{sg}}\, c}{2\,\pi\, r_0}. \tag{3.44}$$

[3] Against the common conventions in this thesis, we use the frequency f_{A} instead of the angular frequency ω_{A} here, as the latter is more intuitive. Both are related by $\omega_{\mathrm{A}} = 2\,\pi\, f_{\mathrm{A}}$.

f_A depends on the grid order, the radius and the sound propagation velocity. The exact structure of the spatial aliasing contributions depends on the distribution of the sampling nodes and the sphere configuration of the array, cf. Section 3.6.1.

In practice, spatial aliasing contributions for $f \leq f_\mathrm{A}$ can be neglected, whereas for $f > f_\mathrm{A}$ we must expect strong aliasing contributions and considerable impairments that impede a reasonable sound field decomposition.

In order to analyze the properties and impacts of spatial aliasing, appropriate signals need to be generated. In order to improve numerical stability, a rigid sphere array configuration is used for the simulations. Eq. (3.24) describes an ideal plane wave,

$$\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega) = 4\,\pi\,\mathrm{i}^n \left[j_n(\tfrac{\omega}{c}\,r_0) - \frac{j_n'(\tfrac{\omega}{c}\,r_0)}{h_n'^{(2)}(\tfrac{\omega}{c}\,r_0)} h_n^{(2)}(\tfrac{\omega}{c}\,r_0) \right] Y_n^m(\theta_\mathrm{w},\phi_\mathrm{w})^*.$$

Spatial sampling is introduced, by using a discrete sampling grid with M_sg nodes located at node angles $(\theta_{g_\mathrm{sg}}, \phi_{g_\mathrm{sg}})$, $g \in [1, M_\mathrm{sg}]$ with corresponding node weights w_{g_sg}:

$$\mathring{P}^{\mathrm{RS\,A}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega) = \sum_{g_\mathrm{sg}=1}^{M_\mathrm{sg}} w_{g_\mathrm{sg}} \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega) \times Y_n^m(\theta_{g_\mathrm{sg}},\phi_{g_\mathrm{sg}}) \right] Y_n^m(\theta_{g_\mathrm{sg}},\phi_{g_\mathrm{sg}})^*. \tag{3.45}$$

$\mathring{P}^{\mathrm{RS\,A}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega)$ denotes the spatial Fourier coefficients including spatial aliasing artifacts due to discrete spatial sampling. The term in square brackets describes an inverse spatial Fourier transform yielding the required sound pressure values at nodes $(\theta_{g_\mathrm{sg}}, \phi_{g_\mathrm{sg}})$.

We define a typical scenario with a Lebedev sampling grid of order $N_\mathrm{sg} = 5$ with $M_\mathrm{sg} = 50$ sampling nodes located on the sphere S_0 defined by the array radius $r_0 = 0.1\,\mathrm{m}$. According to Eq. (3.44), the aliasing frequency is located at around $f_\mathrm{A} \approx 2700\,\mathrm{Hz}$ for this configuration.

A plane wave of full modal order and unit gain with frontal incidence direction ($\theta_\mathrm{w} = \pi/2, \phi_\mathrm{w} = 0$) is generated using Eq. (3.24). Discrete spatial sampling is applied using Eq. (3.45). Analogous to the notation from Eq. (3.21), the corresponding PWD yields

$$D^\mathrm{A}(\theta_\mathrm{d},\phi_\mathrm{d},\omega) = 4\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\,r_0)\, \mathring{P}^{\mathrm{RS\,A}}_{nm\,\mathrm{pw}(\theta_\mathrm{w},\phi_\mathrm{w})}(r_0,\omega)\, Y_n^m(\theta_\mathrm{d},\phi_\mathrm{d}), \tag{3.46}$$

where $D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)$ denotes the decomposition output for a spatially sampled plane wave impact. The decomposition order is set to $N = N_{\mathrm{sg}} = 5$. A scanning PWD along the horizontal plane ($\theta_{\mathrm{d}} = \pi/2, \phi_{\mathrm{d}} = [-\pi, \pi]$) is performed analogous to Figure 3.5. The result of this operation is depicted in Figure 3.40.

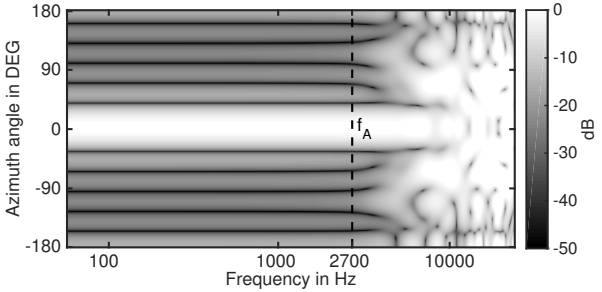


Figure 3.40 Normalized magnitude response $|D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)/(N+1)^2|$ versus the temporal frequency of the PWD (regular beampattern) in the azimuth plane ($\theta_{\mathrm{d}} = \pi/2$, $\phi_{\mathrm{d}} = [-\pi, \pi]$) for a full order broadband plane wave impact from ($\theta = \pi/2$, $\phi = 0$) at the decomposition order $N = 5$.

The impact of spatial aliasing is clearly visible. For $f < f_{\mathrm{A}}$ the PWD performs well and the result is comparable to the one depicted in Figure 3.5. However, at $f = f_{\mathrm{A}}$ the beam starts to burst and for $f \gg f_{\mathrm{A}}$ no reasonable PWD can be performed anymore. The spatial selectivity gets lost and incoming waves cannot be assigned to dedicated directions.

The depicted PWD is normalized by the DI $1/(N+1)^2$, i.e. during ideal operation the output exactly reaches a maximum of 0 dB for $\phi_{\mathrm{d}} = \phi_{\mathrm{w}}$. However, in the aliased frequency range $f \gg f_{\mathrm{A}}$ the output exceeds 0 dB. The dynamic range of Figure 3.40 is limited to a range between -50 dB and 0 dB in order to maintain a comparable scale with Figure 3.5. The increased output level is due to the principle of modal beamforming that consists of constructive summation as well as suppression of certain modal components. The components are weighted with respective radial filters, cf. Section 3.6. Since the orthogonality criterion is violated in the aliased range, the respective mechanisms do not work properly, which leads to increased output levels.

The composite signal $C(\omega)$, introduced in Section 3.4, can be considered an informative control signal. The signal provides an estimator for the overall free field frequency response of the system. Sending a plane wave with unit magnitude from any spatial direction through an ideal system produces an ideal Dirac pulse in the time domain

yielding a perfectly flat magnitude response of 0 dB in the frequency domain at the output.

However, if the signal chain is impaired (e.g. by spatial aliasing), an overall impact on the time-frequency domain can be observed. Thus, $C(\omega)$ is a useful and informative control signal for the assessment of the global spectral impairment due to deficiencies in the system chain.

Under ideal conditions the composite signal $C(\omega)$ yields

$$C(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}}\, D(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega), \tag{3.47}$$

which corresponds to Eq. (3.8). Analogous, we define an aliased composite signal $C^{\mathrm{A}}(\omega)$ that allows for assessing the impact of spatial aliasing on the time-frequency domain yielding

$$C^{\mathrm{A}}(\omega) = \frac{1}{4\pi} \sum_{g_c=1}^{M_{\mathrm{cg}}} w_{g_{\mathrm{cg}}}\, D^{\mathrm{A}}(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}}, \omega). \tag{3.48}$$

An order-matched composite grid with $N_{\mathrm{cg}} = 5$, $M_{\mathrm{sg}} = 50$ grid nodes $(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}})$, and grid weights $w_{g_{\mathrm{cg}}}$ is used for recomposition. $C(\omega)$ and $C^{\mathrm{A}}(\omega)$ are depicted in Figure 3.41.

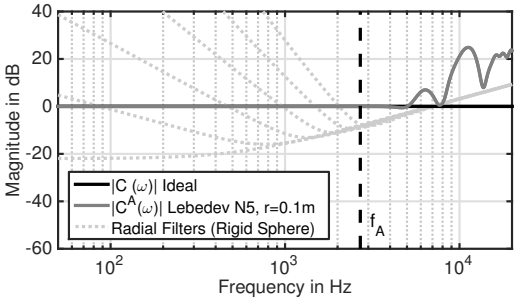


Figure 3.41 Control signals $|C(\omega)|$ and $|C^{\mathrm{A}}(\omega)|$. Additionally, the radial filters for the simulated configuration are depicted. Radial filters are discussed in Section 3.6.

As expected, $|C(\omega)|$ shows a flat magnitude response over the entire frequency range. In contrast, $|C^{\mathrm{A}}(\omega)|$ indicates progressively increased output levels for $f \gg f_{\mathrm{A}}$, similar to the observations in Figure 3.40.

Even if $|C^{\mathrm{A}}(\omega)|$ can be used for roughly assessing global impairments in the overall frequency response of the system, it does not provide specific information on the impairment of the single PWD signals $D(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)$.

In order to gain quantitative information on the impact of spatial aliasing on the PWD signals, the ideal signal is subtracted from the aliased signal and normalized by the DI,

$$A(\omega) = \frac{D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) - D(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)}{(N+1)^2}. \tag{3.49}$$

Hence, the absolute signal contributions that arise due to spatial aliasing for a unit gain plane wave impact are isolated. The latter can be considered as additive spatial noise components that are produced by the system itself. To give a representative overview, $A(\omega)$ is averaged over the entire sphere with a resolution of 1° ($\theta_{\mathrm{d}} = [-\pi, \pi] \times \phi_{\mathrm{d}} = [0, 2\pi[$) yielding the averaged absolute spatial aliasing noise contribution $\bar{A}(\omega)$ for a DI-normalized PWD. The result is depicted in Figure 3.42.

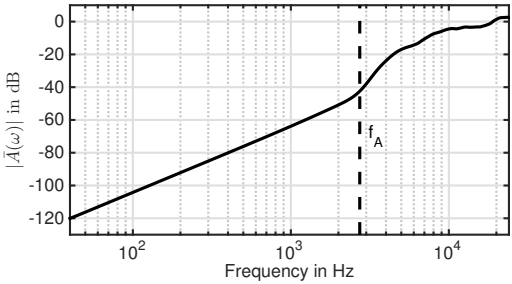


Figure 3.42 Magnitude of the averaged absolute aliasing contribution $|\bar{A}((\omega)|$.

The figure indeed indicates progressively ascending aliasing noise contributions over the entire temporal frequency range as previously discussed. For $f < f_{\mathrm{A}}$ the increase

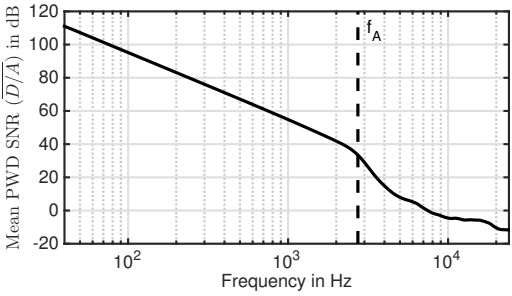


Figure 3.43 Mean SNR $(\overline{D/A})$ of the PWD signals.

is linear on the logarithmic frequency scale and starts to increase disproportionately at $f = f_{\mathrm{A}}$. For very high frequencies, the aliasing noise contributions exceed the maximum

gain of the PWD that is normalized to 0 dB. Referring the ideal PWD signal to the spatial aliasing noise contributions yields the SNR $\overline{D/A}$ of the PWD signals, which is depicted in Figure 3.43.

Analogous, referring $C(\omega)$ to $[C^{\mathrm{A}}(\omega) - C(\omega)]$ yields the overall SNR of the recomposed signal as depicted in Figure 3.44. Generally, the overall SNR is considerably higher compared to the SNR of the single PWD signals, except for $f \gg f_{\mathrm{A}}$. The overall SNR does not account for the spatial distortion but can be used as estimator for the global spectral distortion of the final output signal in the time-frequency domain.

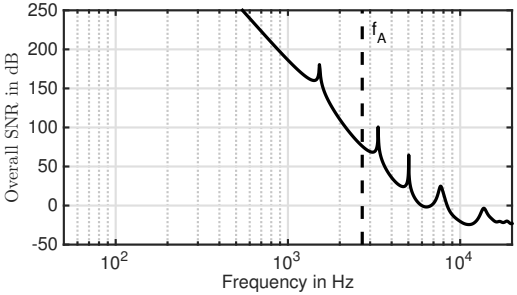


Figure 3.44 Overall SNR of the recomposed signal.

In order to demonstrate the specific impact of spatial aliasing on RHRTFs, both $\mathring{P}_{nm\,\mathrm{pws}(\theta_{\mathrm{w}},\phi_{\mathrm{w}})}(r_0,\omega)$ and $\mathring{P}^{\mathrm{A}}_{nm\,\mathrm{pws}(\theta_{\mathrm{w}},\phi_{\mathrm{w}})}(r_0,\omega)$ are inserted into Eq. (3.11) for comparing clean RHRTFs and RHRTFs containing spatial aliasing artifacts. Both signals are depicted in Figure 3.45 for different wave incidence directions ($\theta_{\mathrm{w}} = \pi/2, \phi_{\mathrm{w}} = \{0, \pi/6, \pi/3, 3\pi/2\}$) . The sampling parameters are chosen according to the example scenario given above. As expected, for $f \leq f_{\mathrm{A}}$ the responses of both signals are nearly identical and spatial aliasing can be neglected. At $f > f_{\mathrm{A}}$ the structures of the aliased RHRTFs differ individually from the clean RHRTFs, which is due to the inherent loss of spatial selectivity. Furthermore, a global increase of high frequency appears. A similar is observed in the control signal $C^{\mathrm{A}}(\omega)$, when the composed PWD signals are not weighted with HRTFs.

Note that this analysis is restricted to our specific example configuration. However, comparable outcomes and dimensions can be found for any other configuration.

More detailed analysis of the properties and inner structures of spatial aliasing can be found e.g. in (Rafaely, 2005), (Rafaely et al., 2007b), (Meyer and Elko, 2008), or (Zotter, 2009a). Analysis of the influences of the sampling node distribution on the spatial distortion and the crosstalk behaviour of aliased higher modes into lower modes are discussed therein.

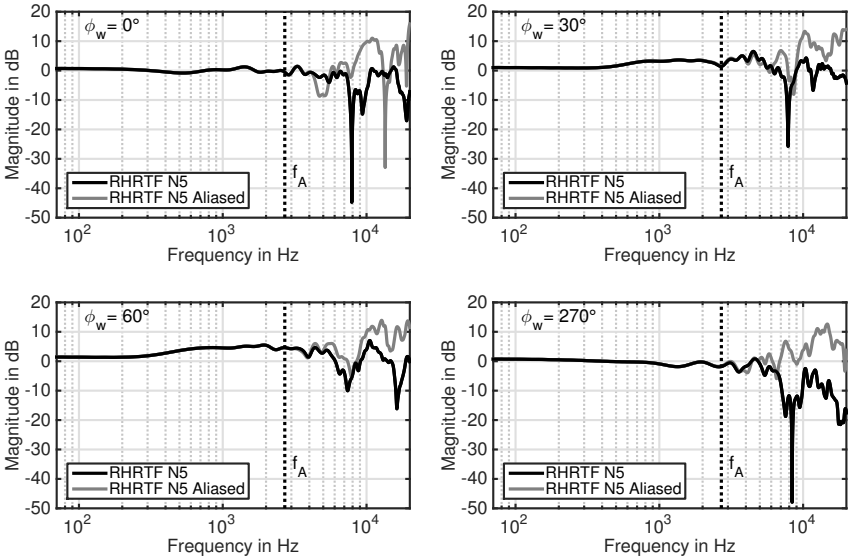


Figure 3.45 Comparison of clean HRTFs and HRTFs containing spatial aliasing artifacts due to discrete spatial sampling for different wave incidence directions ($\theta_\mathrm{w} = \pi/2, \phi_\mathrm{w} = \{0, \pi/6, \pi/3, 3\pi/2\}$).

The inner structure of aliasing is well described in the literature stated above, but it is not very relevant for the considerations in this work. We summarize the important aspects concerning spatial aliasing in practice:

- Plane waves in natural sound fields are not order-limited and thus discrete spatial sampling is always accompanied by (a certain amount of) spatial aliasing.
- Spatial aliasing is present in the entire temporal frequency range. But high temporal frequencies are more prone to spatial aliasing than low frequencies.
- f_A approximately describes a temporal frequency where spatial aliasing artifacts start to surge disproportionately.
- f_A depends on the measurement radius r_0 and the spatial sampling grid order N_sg. Hence, f_A implicitly depends on the node density and grid efficiency.
- For $f \leq f_\mathrm{A}$ spatial aliasing is negligible in practice.
- For $f > f_\mathrm{A}$ strong spatial aliasing artifacts arise that cannot be neglected.

- Spatial aliasing entails spatial distortions. The spatial selectivity of the system gets lost in the affected temporal frequency range and the sound field cannot be decomposed appropriately.
- Spatial aliasing entails distortions of the temporal frequency response for both, the single PWD signals as well as for a recomposed output signal. The system inherently adds spatial noise, which for $f \gg f_\mathrm{A}$ tends to exceed the magnitude of the input signal and leads to increased overall output levels.
- For $f > f_\mathrm{A}$ the response of the RHRTFs is impaired depending on the wave incidence direction. An increased output level can be observed for $f \gg f_\mathrm{A}$.

3.8.1 Reduction of Spatial Aliasing Artifacts

In this section we briefly discuss different options for reducing spatial aliasing artifacts. Different approaches to treat spatial aliasing in the context of spherical microphone arrays are discussed in (Abhayapala et al., 1999), (Rafaely et al., 2007b), (Li and Duraiswami, 2007), (Meyer and Elko, 2008), (Alon and Rafaely, 2012), or (Bernschütz, 2012a).

The most obvious solution for reducing spatial aliasing contributions is either increasing the number of sensors M_sg or decreasing the measurement radius r_0. According to the last section and Eq. (3.44), we assume negligible aliasing contributions for $f \leq (N_\mathrm{sg}\, c)/(2\,\pi\, r_0)$. The number of required nodes, M_sg, for achieving the grid order N_sg can be approximated using Eq. (3.2). Figure 3.46 shows the required number of nodes for fulfilling the relation $f = (N_\mathrm{sg}\, c)/(2\,\pi\, r_0)$ with different sampling grid types and different measurement radii.

Generally, the number of nodes quickly rises with the temporal frequency due to the quadratic relation between N_sg and M_sg. Decreasing the measurement radius is quite efficient. Nevertheless, in Section 3.6.8 we have shown that the array size should not be smaller than the average size of a human head. Hence, the potential for decreasing the radius is limited, unless two concentric arrays with different radii are used. If we exhaust the absolute minimum limit with respect to the radial filters by using an array radius of $r_0 = 9\,\mathrm{cm}$ and assume using an ideal sampling grid with $\eta_\mathrm{g} = 1$, we still need about $M_\mathrm{sg} \approx 1150$ nodes to achieve negligible aliasing contributions up to $f = 20\,\mathrm{kHz}$. Lowering the upper target frequency limit to $f = 10\,\mathrm{kHz}$ and neglecting aliasing for $f > 10\,\mathrm{kHz}$, the amount of nodes reduces to approximately $M_\mathrm{sg} \approx 300$ sensors. Constructing a real-time array with $r_0 = 9\,\mathrm{cm}$ covered with $M_\mathrm{sg} = 300$ discrete microphones is still challenging and expensive in practice. Not only the microphones need to be considered, but also a corresponding amount of microphone preamplifiers

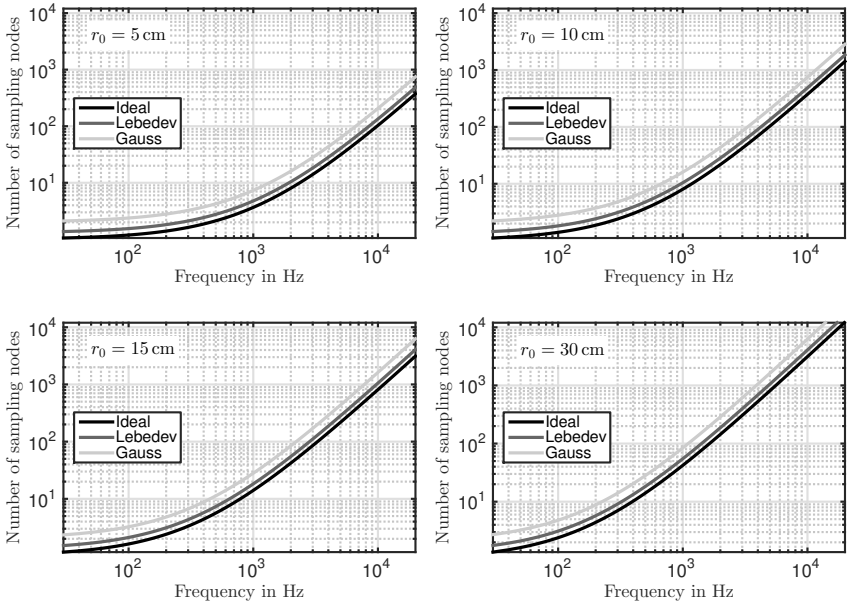


Figure 3.46 Number of required spatial sampling nodes, M_{sg}, versus the temporal frequency for different measurement radii $r_0 = \{5, 10, 15, 30\}\,\mathrm{cm}$. Ideal refers to a grid efficiency $\eta_{\mathrm{g}} = 1$.

and ADCs. A realistic setup could, for instance, involve around 50-120 sensors. Using Lebedev quadratures for sampling, we find constellations with $M_{\mathrm{sg}} = \{50, 74, 86, 110\}$ nodes in the given range. This corresponds with grid orders $N_{\mathrm{sg}} = \{5, 6, 7, 8\}$ and respective aliasing frequencies in the range of approximately 3 kHz to 5 kHz.

A second option for reducing spatial aliasing is to apply spatial anti-aliasing filters as proposed by Rafaely et al. (2007b) and Meyer and Elko (2008). This approach is discussed in the context of expanded transducers in Section 3.10.

Meyer and Elko (2008) propose discarding the array processing in the aliased range and falling back on single microphones pointing approximately towards the incidence direction. The directivity of the microphones or the natural scattering using a rigid-sphere array is used for achieving a certain directivity in this range. This solution does not preserve any closed mathematical description but could be a useful approach in practice.

Alon and Rafaely (2012) propose a theoretical anti-aliasing approach using matrix regularization. However, the approach is not robust against noise and hence not applicable in practice in its current state.

An approach for bandwidth extension based on specific assumptions concerning natural sound fields is proposed in (Bernschütz, 2012a). The method is referred to as bandwidth extension for microphone arrays (BEMA) and discussed in Section 3.11.

Recent approaches to reduce spatial aliasing based on sparse signal representation theory are proposed e.g. by Koyama et al. (2014).

3.8.2 Spectral Compensation

Avoiding spatial aliasing turns out to be quite difficult in practice. The simple alternative is to accept the presence of spatial aliasing. From a purely technical point of view spatial aliasing is obviously not tolerable. The question arises, to what extent spatial aliasing is perceivable by a listener and which kind of perceptual impairments emerge. This question can only be answered by performing listening experiments. Such tests are presented in Section 5.6.3 and Section 5.7.4.

Spatial aliasing creates two different problems. The first is the typical loss of spatial resolution in the aliased frequency range, cf. Figure 3.40. Since we accept aliasing artifacts in this context, we assume that the spatial resolution is irretrievably lost.

The second problem is an increased output level in the aliased frequency range due to additive spatial self-noise. This phenomenon can be observed throughout Figures 3.40, 3.41, 3.44 and 3.45.

Even if the level increase depends on the incidence direction and is not constant for all RHRTFs, a certain amount of increase can be observed for all RHRTFs in common. By averaging the spectral differences between aliased RHRTFs and HRTFs for a representative amount of incidence directions distributed on the sphere, the mean spectral deviation can be estimated. Suitable inversion of the mean spectral deviation yields a global compensation filter for minimizing the average spectral impairment.

The mean spectral deviation between analytic RHRTFs and measured HRTFs using a Lebedev sampling grid with $M_{\mathrm{sg}} = 86$ nodes and grid order $N_{\mathrm{sg}} = 7$ on a measurement radius of $r_0 = 0.0875\,\mathrm{cm}$ with non-critical radial filters is calculated in a first step. Next, the procedure is repeated using measured data, i.e. using measured array based binaural room impulse responses (ABRIRs) and BRIRs from two different rooms, cf. Section 4.3.3. The ABRIRs are measured with an identical array configuration and parameters that are used in the simulations for generating the RHRTFs. Since the measured BRIRs are only available in the horizontal plane, the averaging is restricted

to 360 spatial directions in the horizontal plane. The mean deviations resulting from the analytic simulation, as well as the mean deviations from both measured scenarios are depicted in Figure 3.47 (left plot). A spectral compensation filter is derived from the analytic deviation by magnitude inversion and smoothing. The filter is plotted in Figure 3.47 (right plot).

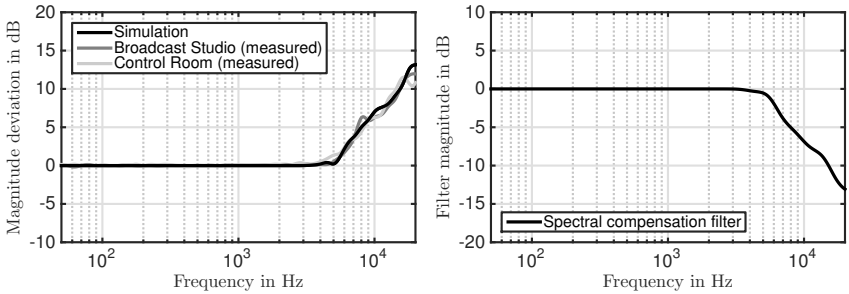


Figure 3.47 Averaged magnitude deviation due to spatial aliasing (left plot) and derived spectral compensation filter (right plot).

The figure shows a remarkable congruence of the simulated prediction and the measured deviations. The increase is linear for $f > f_\mathrm{A}$ and closely matches 6 dB/oct. A compensation filter can be generated by inversion of the magnitude. Owing to the simple structure of the filter slope, a first-order low-pass filter providing the required damping of 6 dB/oct for $f > f_\mathrm{A}$ could be applied as an alternative. Most probably this result that is based on a single example can be generalized. However, this is not further investigated in this thesis.

The derived filter can be applied to both ear signals and it compensates for the high frequency slope induced by spatial aliasing for a specific array setup. Hence, the loss of spatial selectivity due to spatial aliasing still persists, but the global spectral impairment is equalized as far as possible.

The suitability of this approach is perceptually evaluated in Section 5.7.4. The respective stimuli with applied spectral compensation filters have the postfix «EQ» (equalized).

3.9 Effective Operational Bandwidth Including Spatial Aliasing

The effective operational bandwidth (EOB) is introduced in Section 3.6.6, where it is analyzed for a single factor of influence only, i.e. for limiting the radial filter amplification. At this point we pick up both, radial filter limiting at low frequencies and

spatial aliasing at high frequencies. The EOB of the microphone array is substantially narrowed. In order to give a practical example for the dimensions of the resulting EOB, we fall back on the reference array configuration (rigid sphere array with $r_0 = 10\,\mathrm{cm}$ and a Lebedev sampling scheme with $M_{\mathrm{sg}} = 86$ and $N_{\mathrm{sg}} = 7$). Non-critical radial filters according to Section 3.6.5 are used. An analytic full-order plane wave with unit gain arriving from $(\theta_{\mathrm{w}} = \pi/2, \phi_{\mathrm{w}} = 0)$ is generated using Eq. (3.24). Spatial sampling is simulated using Eq. (3.45).

A scanning PWD along the horizontal plane $(\theta_{\mathrm{d}} = \pi/2, \phi_{\mathrm{d}} = [-\pi, \pi])$ is performed using

$$D^{\mathrm{A,L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega) = \frac{4\pi}{(N+1)^2} \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{d}_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0)\, \mathring{P}^{\mathrm{RS\,A}}_{nm\,\mathrm{pw}(\theta_{\mathrm{w}},\phi_{\mathrm{w}})}(r_0, \omega)\, Y_n^m(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}), \tag{3.50}$$

where the decomposition order is set to $N = N_{\mathrm{sg}} = 7$. The magnitude $|D^{\mathrm{A,L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ is depicted in Figure 3.48.

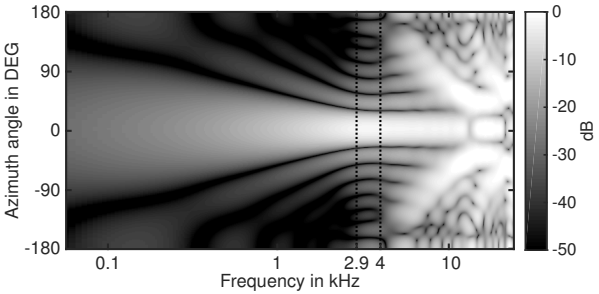


Figure 3.48 Magnitude response $|D^{\mathrm{A,L}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ versus the temporal frequency of the PWD (regular beampattern) in the azimuthal plane $(\theta_{\mathrm{d}} = \pi/2,\ \phi_{\mathrm{d}} = [-\pi, \pi])$ for a full order broadband plane wave impact from $(\theta = \pi/2,\ \phi = 0)$ at the decomposition order $N = 7$.

The plot shows a remaining EOB of approximately 0.5 oct only. The effective operational range is located between roughly 2.9 kHz and 4 kHz. The EOB can be increased at the top end using more sensors or at the bottom end by allowing for additional radial filter gain, cf. Section 3.6.6. Using more sensors in a real-time array is expensive because of the quadratic surge in the number of required sensors. Increasing the radial filter gain amplifies equipment noise, cf. Section 3.7.

Figure 3.49 is generated analogous to Figure 3.48 but with an additional radial filter amplification gain of $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$. The EOB increases to approximately 1 oct. Note

that raising the radial filter amplification gain only sparsely increases the EOB. The bandwidth gain depending on the radial filter amplification limits is quantified in Figure 3.29.

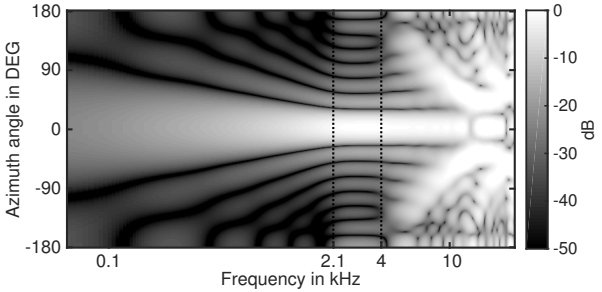


Figure 3.49 Magnitude response analogous to Figure 3.48 but allowing for additional radial filter amplification gain of $\hat{a}_{\mathrm{dB}} = 18\,\mathrm{dB}$.

A nested array with concentric radii can be used for increasing the EOB. The signals from different radii are combined according to their optimum operational ranges. This approach is used in (Melchior et al., 2009) or (Melchior, 2011), for instance. With a scanning array, different radii can be measured sequentially.

Hence, the balance between achievable EOB and technical effort or noise constraints is a question of requirements and budget in practice.

For the purpose of binaural auralization, the comparably low EOB of the single sphere configuration might be acceptable. As discussed before, the lower frequency range with reduced modal resolution is not critical, as long as the array provides a certain minimum size. The aliased high frequency range is more critical. The use of compensation filters for equalizing the spectral impairment that is evoked by spatial aliasing is proposed. However, the spatiotemporal resolution of the array is lost in this range. Even if from a purely technical point of view the signal is not useful, auralization might still work reasonably well. This finally depends on the perceptual impact of the spatial aliasing artifacts (at comparably high frequencies), which is evaluated in listening experiments that are presented in Section 5.6.3 and Section 5.7.4.

3.10 Surface Expansion of the Transducers

For the theoretical consideration of discrete spatial sampling we assumed ideal point transducers with infinitesimal surface expansion on the sphere. In the following, the influence of a transducer covering a certain radial surface with diameter d_{t} on a sphere

of radius r_0 is analyzed. A simplified mathematical model of a spherical cap is used for this purpose. The spherical cap is placed at the north pole around the z-axis, which entails a compact formulation in the spherical wave spectrum domain. In the space domain, the cap is described by

$$T(\theta) = \begin{cases} 1 & \text{for} \quad 0 \leq \theta \leq \gamma_\mathrm{t}/2 \\ 0 & \text{for} \quad \gamma_\mathrm{t}/2 < \theta \leq \pi, \end{cases} \tag{3.51}$$

where θ denotes the elevation angle, counting from the positive z-axis, $\mathrm{T}(\theta)$ the amplitude of the resulting function on the sphere, and γ_t the total aperture angle of the cap, cf. Figure 3.50.

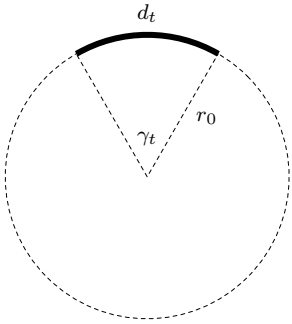


Figure 3.50 Sphere of diameter r_0 (array) with a spherical cap (transducer) at the north pole spanning an arc length d_t and the corresponding angle γ_t.

The aperture angle γ_t describes the ratio between d_t and r_0 yielding

$$\gamma_\mathrm{t}(d_\mathrm{t}) = \frac{d_\mathrm{t}}{r_0}. \tag{3.52}$$

Owing to the established symmetry around the z-axis, Eq. 3.51 can be expressed using a compact expression in the spherical wave spectrum domain (Williams, 1999, p 215) based on Legendre polynomials (Pollow et al., 2012):

$$\mathring{T}_{nm}(\gamma_\mathrm{t}) = \sqrt{\pi\,(2\,n+1)}\;\delta_{m0} \int_{cos\,\frac{\gamma_\mathrm{t}}{2}}^{1} P_n(x)\,dx, \tag{3.53}$$

where P_n denotes the Legendre polynomial of order n and δ_{m0} the Kronecker delta. $\mathring{T}_{nm}(\gamma_\mathrm{t})$ are the spatial Fourier coefficients, which are zero for $m \neq 0$ and entirely

real in the given case. The cap at the north pole can be rotated in the spherical wave spectrum domain to an arbitrary position on the sphere using Wigner-D rotation, cf. Section 2.16.1. Applying the inverse spatial Fourier transform to $\mathring{T}_{nm}(\gamma_\mathrm{t})$ yields $\mathrm{T}(\theta)$. To be exact, it yields $\mathrm{T}(\theta,\ \phi)$, since the function is defined over the entire sphere and not restricted to a circle. An example for the inverse transform of $\mathring{T}_{nm}(\gamma_\mathrm{t})$ for an angle $\gamma_\mathrm{t} = pi/2$ is depicted in Figure 3.51. For generating the image, inverse spatial Fourier transforms are applied for 360 × 181 angles covering the sphere. The magnitudes are mapped to a flat surface according to a simple cylindrical Plate Carrée projection (Bernschütz, 2012b). The transform is truncated at $N = 20$, which entails a soft edge at $\theta = \pi/12$ (15°) and some visible ripple in the space-frequency domain.

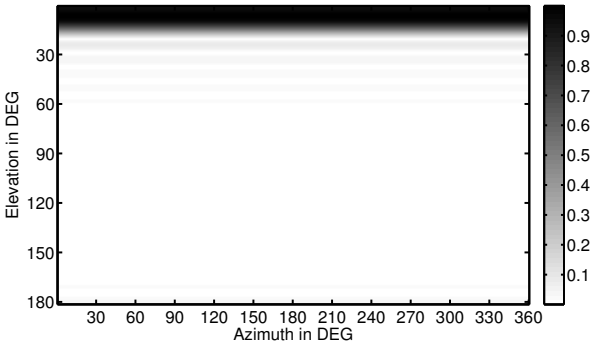


Figure 3.51 360 x 181 point normalized inverse spatial Fourier transform of $\mathring{T}_{nm}(\gamma_\mathrm{t})$ for $\gamma_\mathrm{t} = \pi/6$ truncated at $N = 20$.

The modal signal power p_n that corresponds to the spectrum $\mathring{T}_{nm}(\gamma_\mathrm{t})$ for a specific spherical harmonic order n can be calculated using (Pollow et al., 2012)

$$p_n = \sum_{m=-n}^{n} |\mathring{T}_{nm}(\gamma_\mathrm{t})|^2. \tag{3.54}$$

p_n remains constant when rotating the spherical cap using Wigner-D functions (Pollow et al., 2012). Observing the absolute modal signal power p_n of the spherical harmonic spectrum $\mathring{T}_{nm}(\gamma_\mathrm{t})$ is not too meaningful at this point. However, relating the latter to the modal signal power that is excited by a infinitesimal point would show the impact of surface expansion on the modal signal power in the spherical harmonic spectrum referred to ideal conditions.

The expanded spherical cap described by Eq. 3.53 is reduced to an infinitesimal point at the north pole for $\gamma_\mathrm{t} \to 0$ yielding

$$\mathring{T}'_{nm} = \lim_{\gamma_\mathrm{t} \to 0} \sqrt{\pi\,(2\,n+1)}\,\delta_{m0} \int_{cos\,\frac{\gamma_\mathrm{t}}{2}}^{1} P_n(x)\,dx, \tag{3.55}$$

with its corresponding signal power p'_n analogous to Eq. 3.56

$$p'_n = \sum_{m=-n}^{n} |\mathring{T}'_{nm}(\gamma_\mathrm{t})|^2. \tag{3.56}$$

The ratio p_n/p'_n describes the relative modal signal power of the spherical harmonic spectrum excited by a function describing a spherical cap with expansion d_t related to an infinitesimal point. Applying reciprocity to the analytic description, a transducer in the array that is exposed to a spherical harmonic mode of order n, would deliver an output power according to p_n. From this point of view, p_n/p'_n describes the relative output power of a transducer with diameter d_t referred to the ideal point transducer, when exposed to spherical harmonic modes of order n.

Figure 3.52 shows an example for the relative output power $10\log_{10}(p_n/p'_n)$ in dB for $\gamma_\mathrm{t} = d_\mathrm{t}/r_0 = \pi/6$ plotted as a function of the spherical harmonic order $n = [0, 40]$. The plot reveals a sinc function. The spherical cap defined in Eq. 3.51 describes a rect-function on the sphere in the space(-frequency) domain. Given the well-known correspondences of the akin time-frequency Fourier transform, cf. e.g. (Ohm and Lüke, 2004, p 62), the resulting sinc function in the spherical wave spectrum domain is hardly surprising.

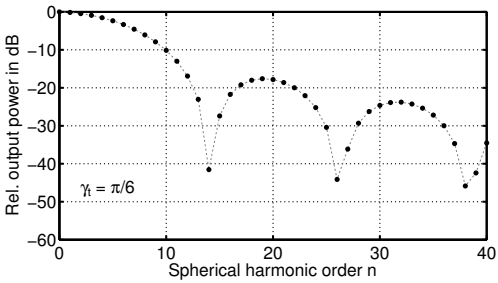


Figure 3.52 Relative transducer output power for $\gamma_\mathrm{t} = \pi/6$ referred to an ideal point transducer at spherical harmonic orders $n = [0, 40]$.

Depending on the point of view, this outcome indicates certain drawbacks of using expanded transducers in an ideal system on the one hand, or, it indicates the chance

to take advantage of the resulting power loss for realizing basic spatial low-pass filters on the other hand. Both perspectives are discussed in the following.

3.10.1 Expanded Transducers in Ideal Systems

From the perspective of an ideal system, the question arises, to what extent expanded transducers would impair the system properties. The increasing power loss at higher modal orders observed in Figure 3.52 would generally impair the PWD or any desired beam pattern to a certain extent, as long as the power loss is not compensated. A much more critical problem are the complete drops in output signal power of the expanded transducer. At certain orders, e.g. $n = 14$ in the previous example, the transducers are not able to deliver a reasonable output signal power. Compensating these drops would be highly unstable and amplify the transducer noise excessively. As a consequence, compensation is not feasible in practice.

Due to the excessive radial filter amplification, a microphone array is generally not able to resolve higher spherical harmonic orders. The previous example with $\gamma_\mathrm{t} = \pi/6$ is exaggerated for microphone arrays, where usually transducers are used whose radius is small compared to the array radius. For spherical loudspeaker arrays, in contrast, the given ratio might be more realistic, e.g. (Pollow et al., 2012). Figure 3.53 shows the relative output power for more realistic examples of a microphone array with a radius of $r_0 = 8.75\,\mathrm{cm}$ (approx. head diameter) and typical transducer diameters of $d_\mathrm{t} = \{1/4\text{”}, 1/2\text{”}, 1\text{”}\}$. Results for an exaggerated transducer diameter of $d_\mathrm{t} = 2\text{”}$ are shown as well.

The plots show a very moderate power loss for the two smallest transducers. The first null of the sinc-function is not even located in the depicted range. For larger transducers, the sinc-function becomes visible again. In this thesis, the measured array data is acquired using an array of $r_0 = 8.75\,\mathrm{cm}$ that is equipped with a small transducer, $d_\mathrm{t} < 1/4\text{”}$ (Earthworks M30). Furthermore, the measured data is only decomposed up to a maximum order of $N = 7$. Thus, the influence of the transducer expansion is negligible in this context.

The previous studies of the transducer size consider the space domain and the spherical wave spectrum domain only. For large transducers, particularly when using single large diaphragms, resonances and partial oscillations might arise at wave lengths that are small when compared to the transducer diameter. This influences the transmission properties in the time-frequency domain, which is a common problem in microphone design and is not further regarded in this work.

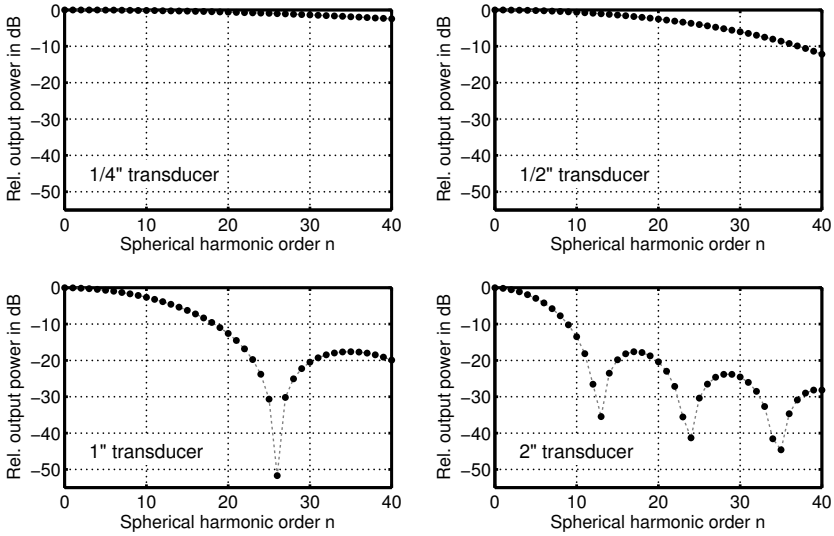


Figure 3.53 Relative transducer output power referred to an ideal point transducer at spherical harmonic orders $n = [0, 40]$ for different diaphragm diameters $d_\mathrm{t} = \{1/4", 1/2", 1", 2"\}$ located on an array radius of $r_0 = 8.75\,\mathrm{cm}$.

3.10.2 Expanded Transducers as Modal Low-pass Filter

Besides causing unwanted impairments to an ideal system, the damping of the output power at higher modal orders due to the transducer expansion can be employed to realize basic modal low-pass filters, refer to Rafaely et al. (2007b) or Meyer and Elko (2008). Modal low-pass filters minimize spatial aliasing artifacts that arise due to discrete spatial sampling of the sound field. The ratio $\gamma_\mathrm{t} = d_\mathrm{t}/r_0$ can be chosen to match certain criteria, e.g. to provide a certain amount of power damping at a specific target mode. Following common conventions, the power drop can be specified to match $-3\,\mathrm{dB}$ at a desired target mode. Figure 3.54 (left plot) shows a modal low-pass filter with target mode $n = 10$ that is obtained for $\gamma_\mathrm{t} \approx 0.31$. Analogous to the given example, low-pass filters for reasonable target orders can be designed by varying the ratio $\gamma_\mathrm{t} = d_\mathrm{t}/r_0$. The ratio γ_t for specific target orders $n = [5, 40]$ is depicted in Figure 3.54 (right plot).

For several reasons a single transducer with large diaphragm is not feasible in practice. In order to realize a spatial low-pass filter, a dense transducer sub-array consisting of small transducers can be used to cover the required area with diameter d_t instead of

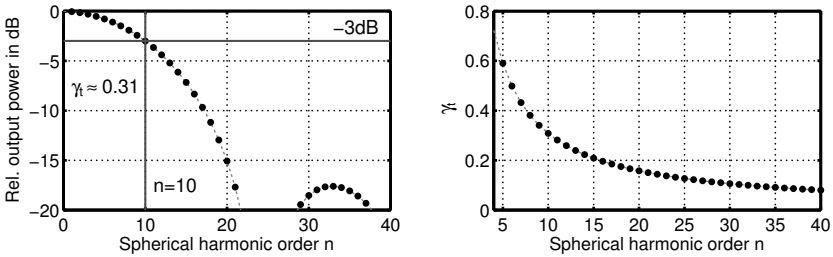


Figure 3.54 Exemplary basic modal low-pass filter for target order $n = 10$ (left plot). Required ratio $\gamma_\mathrm{t} = d_\mathrm{t}/r_0$ for equivalent low-pass filters at target orders $n = [5, 40]$ (right plot).

using a single large transducer. The transducers of the sub-array can be combined by a simple parallel circuit even before the microphone amplification and analog-digital conversion. The sub-array would then be treated as a single transducer of the main array.

Even though the previously discussed basic modal low pass filter is easy to realize and therefore convenient in practice, the resulting filter characteristic in the spherical wave spectrum domain is not satisfying, since the filter is based on a simple rect-function in the space domain. More graceful filters concerning the slew rate, pass-band flatness, and rejection-band damping could be designed in theory. In return, a complex and sophisticated sub-array pattern with individual weighting of the transducers would be required to replace each single transducer position of the main array. This approach is challenging and expensive in practice. The theoretical design of an ideal spatial low-pass filter (rect and sinc functions are exchanged between the space domain and the spherical wave spectrum domain) is discussed in (Rafaely et al., 2007b). Thus, the simple approach of using an expanded transducer (or passive transducer sub-array) that covers a spherical cap could be the most feasible solution for the practical realization of modal low-pass filters, even though the resulting filter properties might not be the best imaginable.

3.11 Bandwidth Extension for Microphone Arrays (BEMA)

In this section we discuss an approach for extending the bandwidth of microphone arrays[4] presented in (Bernschütz, 2012a). In this thesis we only discuss a simplified basic version of the approach in order to outline the underlying idea. The BEMA

[4] The approach is referred to as bandwidth extension for microphone arrays (BEMA).

method can be used to patch spatial aliasing. The underlying approach can also be employed for other applications, like the design of spatial audio data reduction algorithms.

The approach and mathematical description for decomposing the sound field as it is discussed in this thesis and in the common literature that this thesis is based upon, is highly abstract and does not account for potential prior knowledge on the specific conditions in realistic sound fields. This refers to the capability of independently resolving monofrequent waves of arbitrary temporal frequency in both, the time-frequency domain and the space-spherical wave spectrum domain. This might not be necessary under certain assumptions and conditions.

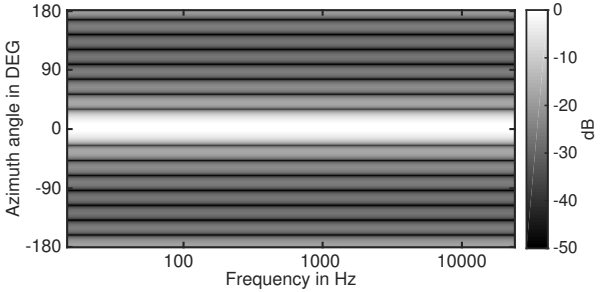


Figure 3.55 Circular response of an ideal array in the azimuthal plane ($\theta_\mathrm{d} = \pi/2$, $\phi_\mathrm{d} = [-\pi, \pi]$) for a broadband plane wave impact from ($\theta_\mathrm{w} = \pi/2$, $\phi_\mathrm{w} = 0$) at decomposition order $N = 7$.

The azimuthal response of an ideal[5] array to a broadband plane wave impact is depicted in Figure 3.55. The response is uniform over the full time-frequency spectrum due to the native constant directivity (CD) property.
As a consequence, we only need to evaluate the decomposition for a single temporal frequency (bin) in order to derive the entire response in the space-spherical wave spectrum domain over the full temporal frequency range. In other words, the spatial direction of a single ideal broadband plane wave impact is sufficiently determined by evaluating a monofrequent portion at arbitrary single temporal frequency.

The effective operational bandwidth (EOB) of a realistic array under typical technical constraints, i.e. involving discrete spatial sampling and radial filter amplification limiting is shown in Section 3.6.6. We can achieve a nearly ideal array response for a

[5] Ideal means not accounting for neither discrete spatial sampling nor radial filter amplification limiting.

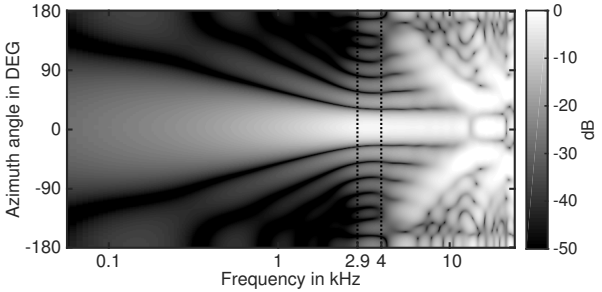


Figure 3.56 Circular response of a non-ideal array with $r_0 = 10\,\mathrm{cm}$, $M_{\mathrm{sg}} = 86$ discrete sampling positions and $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ in the azimuthal plane ($\theta_{\mathrm{d}} = \pi/2$, $\phi_{\mathrm{d}} = [-\pi, \pi]$) for a broadband plane wave impact from ($\theta_{\mathrm{w}} = \pi/2$, $\phi_{\mathrm{w}} = 0$) at decomposition order $N = 7$.

narrow frequency band below $\omega = N\,c/r_0$. The respective band covers the range between approximately 2.9 kHz and 4 kHz for the configuration used to generate Figure 3.56.

Since we found a monofrequent portion to be sufficient to determine the spatial response to a single ideal broadband plane wave on the entire time-frequency spectrum, the stable narrow frequency band indeed delivers sufficient information to perfectly recover the ideal response over the entire time-frequency spectrum. We extract the reference response at $\omega_i = N\,c/r_0$ or slightly below (i.e. around 3.5 kHz in the upper example), since the array response is most robust in this specific range. This narrow-band extract is referred to as spatial image in the following. Once the spatial image is extracted, it can be copied and pasted to the impaired sections in the response. As a consequence, we are able to reconstruct the result from Figure 3.55 based on the narrow-band spatial image extracted from Figure 3.56.

In order to put this into a more concrete mathematical description, we assume arbitrary modal sound field description $\mathring{P}_{nm}(r_0, \omega)$. In a first step we extract a spatial image $\dot{I}'_{nm}$ at a stable temporal frequency ω_{i} using

$$\dot{I}'_{nm} = \mathring{P}_{nm}(r_0, \omega_i)\, d_n(\tfrac{\omega_i}{c}\, r_0), \tag{3.57}$$

where $\omega_{\mathrm{i}} = N\,c/r_0$ is a reasonable choice in this context due to the high robustness of the array response. The extracted spatial image $\dot{I}'_{nm}$ is unchained from its specific dependence on the measurement radius, the array configuration, and the temporal frequency by multiplying the source coefficients $\mathring{P}_{nm}(r_0, \omega_i)$ with the specific radial filters

$d_n(\frac{\omega_i}{c}\, r_0)$, cf. Section 3.6. Thus, $\dot{I}'_{nm}$ represents abstracted spatiotemporal properties of the sound field described in $\mathring{P}_{nm}(r_0, \omega)$ at $\omega = \omega_i$.

In a next step, spatial Fourier coefficients $\mathring{P}^{\mathrm{B}'}_{nm}(r_0, \omega)$ for an arbitrary temporal frequency ω can be generated using

$$\mathring{P}^{\mathrm{B}'}_{nm}(r_0, \omega) = \frac{\dot{I}'_{nm}}{d_n(\frac{\omega}{c}\, r_0)}. \tag{3.58}$$

Hence, we create valid spatial Fourier coefficients $\mathring{P}^{\mathrm{B}'}_{nm}(r_0, \omega)$ for an arbitrary frequency ω that provide (ideal) spatiotemporal properties that are identical with the properties we find in $\mathring{P}_{nm}(r_0, \omega)$ at $\omega = \omega_i$. So far we are able to recover the ideal spatial Fourier coefficients for any ω merely based on the information gained at ω_i. This is strictly limited to ideal plane waves in $\mathring{P}_{nm}(r_0, \omega)$ that actually provide a perfectly flat spectral magnitude response and a linear phase response, since we completely loose or ignore specific magnitude and phase information.

The first stage so far only considers the spatiotemporal properties and neglects any spectrotemporal properties. As a consequence, it cannot be used for practical applications. In real sound fields there are no plane waves that strictly fulfill the ideal properties. Real sound sources do not produce ideal plane waves, as this would imply reproducing an ideal spectrotemporal Dirac pulse. For simplicity, we use (room) impulse responses instead of complex audio signals. The sound source provides a magnitude and phase response that we would like to include into our considerations. If we perform room impulse response measurements, the incident reflections can be considered to be delayed copies of the source signal, modified with respect to magnitude and phase, which precisely characterize the room impulse response. Hence, we need to account for the specific spectrotemporal properties.

At this point we come to the core of the BEMA approach. It consists of merging the abstract spatiotemporal properties that are gained at ω_i with spectrotemporal information that is acquired using a separate omni-directional transducer at the origin of the array. The respective space-frequency domain signal is denoted as $C_0(\omega)$.

We reformulate Eq. (3.57) reading

$$\dot{I}_{nm} = \frac{\mathring{P}_{nm}(r_0, \omega_i)\, d_n(\frac{\omega_i}{c}\, r_0)}{|C_0(\omega_i)|}, \tag{3.59}$$

where the spatiotemporal image is normalized by the magnitude of the center transducer signal $C(\omega)$ at $\omega = \omega_i$. We reformulate Eq. (3.58) in order to account for the spectrotemporal information acquired at the center transducer yielding

$$\mathring{P}^{\mathrm{B}}_{nm}(r_0, \omega) = \frac{\dot{I}_{nm}}{d_n(\frac{\omega}{c}\, r_0)}\, C_0(\omega). \tag{3.60}$$

Hence, we separately account for a spatiotemporal component that describes the spatial distribution unchained from specific spectrotemporal information, and a spectrotemporal component that comprises magnitude and phase information but does not carry specific spatiotemporal information due to the omni-directional characteristics. We mark both separate components in Eq. (3.60),

$$\mathring{P}^{\mathrm{B}}_{nm}(r_0, \omega) = \underbrace{\frac{\dot{I}_{nm}}{d_n(\frac{\omega}{c}\, r_0)}}_{\mathrm{A}} \overbrace{C_0(\omega)}^{\mathrm{B}}.$$

A: Spatiotemporal component: Spatial distribution
B: Spectrotemporal component: Magnitude and phase

Acquiring the spectrotemporal information with a separate transducer at the mathematical and physical origin of the array is particularly suitable, since this spatial position is the reference point of phase in the original Fourier coefficients. Due to the magnitude normalization of the spatial image to $|C(\omega_i)|$ in Eq. (3.59) we automatically achieve suitable magnitude relations. Hence, there is no specific need for calibrating the gain of the center transducer.

We have found a closed form description with magnitude and phase relations that perfectly match the original Fourier coefficients. This is of great relevance for fading between original Fourier coefficients and BEMA Fourier coefficients. We can fade or switch from the original coefficients to BEMA estimated coefficients for patching unstable or impaired sections of the original response without even producing gaps in the magnitude and phase response. However, in practical applications, particularly when using a rigid sphere array, the center transducer cannot be placed at the true origin. The center transducer is then placed on top, below or besides the array. The resulting phase gap may be negligible in practice.

We see that the BEMA approach produces new valid and mathematically closed spatial Fourier coefficients that are fully compatible with the original Fourier coefficients.

3.11.1 BEMA Requirements and Restrictions

Even though the BEMA estimated Fourier coefficients are fully compatible with the original coefficients, they differ from them and there are specific restrictions. With BEMA coefficients there is no true source separation in a classical sense, but the signal $C_0(\omega)$ at the omni-directional transducer is reasonably distributed and assigned in proper portions to dedicated spatial incidence directions. Single separate plane wave incidences can indeed be reconstructed perfectly. Nevertheless, for superimposed plane waves with different magnitude and phase responses, the omnidirectional sum response is distributed and assigned in proper portions to the true directions of incidence. Depending on the scenario, various inherent estimation errors will emerge. But we still achieve substantial improvements e.g. when patching aliased sections of the array response.

BEMA works only for incident waves that provide energy at ω_i, since this is where the spatial image is extracted. If there is no energy at ω_i, the spatial image extraction fails and the wave becomes invisible to the BEMA processing. Hence, there are severe restrictions, e.g. for monochromatic waves with $\omega \neq \omega_i$. However, most waves in realistic sound fields cover sufficient spectral bandwidth. If we restrict our considerations to room impulse responses, the incident waves consisting of direct sound and room reflections usually cover sufficient spectral bandwidth to be fully compatible with BEMA. Room resonances at low frequencies that due to their monochromatic nature would be invisible for the BEMA processing are an exception.

BEMA should not be applied to the full time-frequency bandwidth, but rather be used to patch dedicated impaired sections of the array response only. The approach is well suitable in the context of auralization for patching the upper frequency band that is impaired due to spatial aliasing. Here we further use the original Fourier coefficients $\mathring{P}_{nm}(r_0, \omega)$ for $f < f_\mathrm{A}$ and fade or switch over to the estimated BEMA Fourier coefficients $\mathring{P}^\mathrm{B}_{nm}(r_0, \omega)$ for $f \geq f_\mathrm{A}$. Thus, below the aliasing frequency we have the native array resolution and source separation capabilities and above the aliasing frequency the signal is complemented by a suitable BEMA estimation.

The BEMA processing should be applied in blocks (e.g. 128 or 256 samples with or without overlap) in order to properly resolve complex dynamic structures that vary with time in a measured or simulated room impulse response. Straight block based BEMA processing apparently works very well for dedicated single or superimposed reflections, cf. Section 5.6.4, but it turns out to be rather sensitive to highly diffuse portions in the sound field. Audible block processing artifacts arise in diffuse sound fields, cf. Section 5.7.4.4.

3.11.2 BEMA Spatial Anti Aliasing

In this section specific properties of the BEMA approach are demonstrated. We focus on using BEMA for patching the upper spectrotemporal range $f > f_\mathrm{A}$ that is impaired by spatial aliasing, as proposed in (Bernschütz, 2012a) and evaluated in the listening experiments in Section 5.

Eq. (3.45) delivers spatial Fourier coefficients according to a single plane wave impact when factoring in discrete spatial sampling that results in spatial aliasing artifacts for $f > f_\mathrm{A}$. The equation is modified in order to allow for generating K plane waves with individual incidence directions $(\theta_{\mathrm{w}_k}, \phi_{\mathrm{w}_k})$, amplitudes $\hat{p}_k$, and time shifts τ_k. The modified equation reads

$$\mathring{P}^{\mathrm{RS\,A}}_{nm}(r_0,\omega) = \sum_{g_\mathrm{sg}=1}^{M_\mathrm{sg}} w_{g_\mathrm{sg}} \sum_{k=1}^{K} \hat{p}_k \, \mathrm{e}^{-\mathrm{i}\,\omega\,\tau_k} \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_{\mathrm{w}_k},\phi_{\mathrm{w}_k})}(r_0,\omega) \times Y_n^m(\theta_{g_\mathrm{sg}},\phi_{g_\mathrm{sg}}) \right] Y_n^m(\theta_{g_\mathrm{sg}},\phi_{g_\mathrm{sg}})^*, \quad (3.61)$$

where the coefficients $\mathring{P}^{\mathrm{RS}}_{nm\,\mathrm{pw}(\theta_{\mathrm{w}_k},\phi_{\mathrm{w}_k})}(r_0,\omega)$ are defined in Eq. (3.24). They describe an ideal plane wave impact.

We require an appropriate center transducer signal $C(\omega)$ that reproduces the sum signal of the K superimposed waves. Since the transducer is located at the mathematical origin, the description reduces to the sum of individual amplitude and phase shift terms,

$$C_0(\omega) = \sum_{u=1}^{U} \hat{p}_k \, \mathrm{e}^{-\mathrm{i}\,\omega\,\tau_k}. \quad (3.62)$$

The spatial image $\dot{I}_{nm}$ is extracted using Eq. (3.59) yielding

$$\dot{I}_{nm} = \frac{\mathring{P}^{\mathrm{RS\,A}}_{nm}(r_0,\omega_i)\, d_n(\frac{\omega_i}{c} r_0)}{|C_0(\omega_i)|}, \quad (3.63)$$

where we chose $\omega_\mathrm{i} = (N-1)\,c/r_0$, i.e. slightly below $\omega_\mathrm{i} = N\,c/r_0$ for providing some headroom with respect to the aliasing frequency. The BEMA Fourier coefficients $\mathring{P}^{\mathrm{B}}_{nm}(r_0,\omega)$ are generated using

$$\mathring{P}^{\mathrm{B}}_{nm}(r_0,\omega) = \frac{\dot{I}_{nm}}{d_n(\frac{\omega}{c} r_0)}\, C_0(\omega). \quad (3.64)$$

At this point we have the original Fourier coefficients $\mathring{P}^{\mathrm{A}}_{nm}$ and the BEMA estimated Fourier coefficients $\mathring{P}^{\mathrm{B}}_{nm}$. A patched coefficient set $\mathring{P}^{\mathrm{AB}}_{nm}$ is defined by

$$\mathring{P}^{\mathrm{AB}}_{nm}(r_0,\omega) = \begin{cases} \mathring{P}^{\mathrm{A}}_{nm}(r_0,\omega) & \text{for } f < f_{\mathrm{A}} \text{ and} \\ \mathring{P}^{\mathrm{B}}_{nm}(r_0,\omega) & \text{for } f \geq f_{\mathrm{A}}. \end{cases} \tag{3.65}$$

Hence, at $f = f_{\mathrm{A}}$ we switch over[6] from the original coefficients to the BEMA coefficients. For evaluating the result of this operation, a DI-normalized plane wave decomposition is performed, based on the original coefficients $\mathring{P}^{\mathrm{A}}_{nm}$ yielding

$$D^{\mathrm{A}}(\theta_{\mathrm{d}},\phi_{\mathrm{d}},\omega) = \frac{4\pi}{(N+1)^2} \sum_{n=0}^{N} \sum_{m=-n}^{n} d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0)\, \mathring{P}^{\mathrm{A}}_{nm}(r_0,\omega)\, Y_n^m(\theta_{\mathrm{d}},\phi_{\mathrm{d}}), \tag{3.66}$$

as well as on the patched coefficient set $\mathring{P}^{\mathrm{AB}}_{nm}$ yielding

$$D^{\mathrm{AB}}(\theta_{\mathrm{d}},\phi_{\mathrm{d}},\omega) = \frac{4\pi}{(N+1)^2} \sum_{n=0}^{N} \sum_{m=-n}^{n} d_n^{\mathrm{RS}}(\tfrac{\omega}{c}\, r_0)\, \mathring{P}^{\mathrm{AB}}_{nm}(r_0,\omega)\, Y_n^m(\theta_{\mathrm{d}},\phi_{\mathrm{d}}). \tag{3.67}$$

Three different scenarios (BS1, BS2 and BS3) are analyzed to demonstrate some of the properties of BEMA in the following. The simulated array corresponds to the reference configuration. It is a rigid sphere array with $r_0 = 10\,\mathrm{cm}$ and $M_{\mathrm{sg}} = 86$. The decomposition order is set to $N = 7$ and no radial filter amplification limiting is applied. Other simulation parameters that depend on the specific scenario are listed in Table 3.1.

Table 3.1 Simulation parameters for scenarios BS1, BS2 and BS3.

Scenario	BS1	BS2	BS3
Number of waves K	1	3	3
Direction $(\theta_{\mathrm{w}_k}, \phi_{\mathrm{w}_k})$	$(\pi/2, 0)$	$(\pi/2, 0)$, $(\pi/2, \pi/2)$, $(\pi/2, -\pi/3)$	$(\pi/2, 0)$ $(\pi/2, \pi/2)$, $(\pi/2, -\pi/3)$
Amplitude $\hat{p}_k$	0 dB	0 dB, -6 dB, -12 dB	0 dB, -6 dB, -12 dB
Time shift τ_k	0 ms	0 ms, 0 ms, 0 ms	0 ms, 30 ms, 10 ms

[6] In practical implementations we should fade between both for about 1/6 oct to 1/3 oct in order to achieve a smooth transition.

The following plots show both magnitude responses $|D^{\mathrm{A}}|$ and $|D^{\mathrm{AB}}|$ using decomposition angles ($\theta_{\mathrm{d}} = \pi/2$, $\phi_{\mathrm{d}} = [-\pi, \pi]$) as a function of the temporal frequency.

The first scenario (BS1) is depicted in Figure 3.57. Note that an isolated plane wave impact can be reconstructed perfectly. This still holds true in case the incident wave provides a particular magnitude and phase response. The only crucial condition is that the wave provides energy at $\omega = \omega_i$ for enabling the extraction of the spatial image.

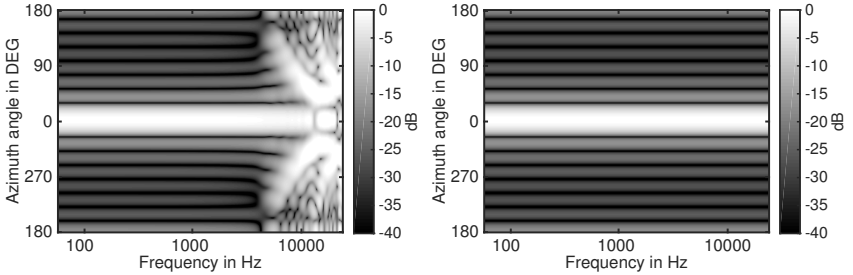


Figure 3.57 Responses $|D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (left) and $|D^{\mathrm{AB}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (right) for scenario BS1.

The second scenario (BS2) is depicted in Figure 3.58. There are three incident waves with different directions and different amplitudes but equal time shifts. We observe that BEMA apparently is also capable of resolving this more complex structure. The response is properly extended and the different magnitudes of the waves are matched. However, we must keep in mind, that above $f \geq f_{\mathrm{A}}$ there is no true source separation possible anymore. The spatial image is applied and the superimposed signal from the center transducer is simply distributed in proper portions to the incident waves.

The third scenario (BS3) is depicted in Figure 3.59. Analogous to BS2 there are three incident waves with different directions and different amplitudes. However, in contrast to BS2, there are different time shifts for each wave. Now we observe the overlay of an irregular pattern in the range of the BEMA estimated coefficients. The pattern is due to comb filter effects that arise at the omni-directional center transducer. At the center transducer all waves are superimposed without directional separation. Hence, BEMA is not capable of reconstructing an ideal response. However, the result is still considerably improved when comparing to the aliased response.

In this thesis we only discuss aspects that are useful for understanding the functional principle of BEMA. Refer to (Bernschütz, 2012a) for a more detailed discussion.

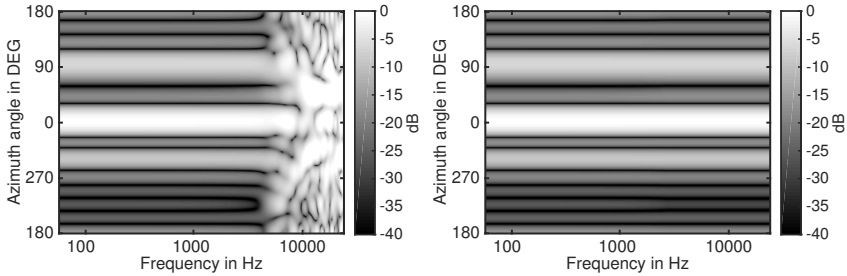


Figure 3.58 Responses $|D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (left) and $|D^{\mathrm{AB}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (right) for scenario BS2.

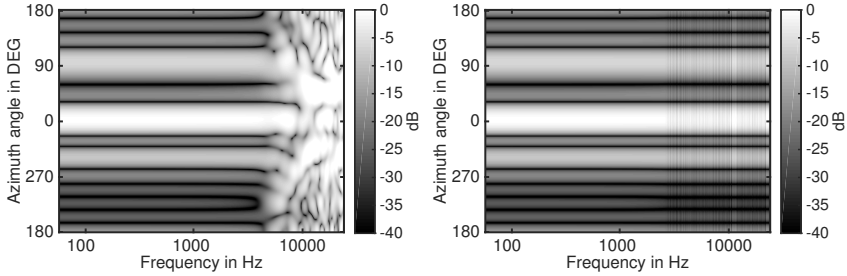


Figure 3.59 Responses $|D^{\mathrm{A}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (left) and $|D^{\mathrm{AB}}(\theta_{\mathrm{d}}, \phi_{\mathrm{d}}, \omega)|$ (right) for scenario BS3.

As mentioned in the introduction of the chapter, the functional principle can be employed for other purposes besides patching aliased portions in the array response. The separation and separate treatment of spectrotemporal and spatiotemporal properties of the sound field is, for instance, a potential approach for designing spatial audio codecs.

3.12 HRTFs

This section shortly discusses some of the most relevant factors, approaches and issues concerning head-related transfer functions (HRTFs) in practice.

3.12.1 Non-individual, Individualized and Individual HRTFs

We distinguish between non-individual, individualized, and individual HRTFs. Non-individual HRTFs are measured with a dummy head or a reference subject. No adapta-

tion is made to the specific listener. Individualized HRTFs are non-individual HRTFs that are further adapted to the specific anatomy of the listener. Individual HRTFs are individually measured or modeled for each specific listener.

Generally, the use of individual HRTFs is preferable, since there is considerable anatomical variation between different subjects or dummy heads. If the mismatch between the subject's anatomy and non-individual HRTFs is too large, the localization accuracy can be affected (Plenge, 1974), (Wenzel et al., 1993), (Møller et al., 1996). Nevertheless, it is not always feasible to use individual HRTF, since the measurement process for human subjects is quite elaborate. Different approaches for individualizing non-individual HRTFs, e.g. by using ITD-scaling, are proposed by Middlebrooks (1999), Zotkin et al. (2003), Hu et al. (2006), or Lindau et al. (2010). As an alternative to active individualization, a best fitting set can be chosen (Seeber and Fastl, 2003), (Katz and Parseihian, 2012) from a database of different HRTFs, such as the CIPIC database by Algazi et al. (2001b), for example.

For the present work, a Neumann KU 100 dummy head was used to acquire the necessary HRTFs and BRIRs. No individual HRTFs are involved nor is any approach to individualization applied for different reasons. Measuring adequate individual HRTF sets was not feasible due to the demands concerning grid resolution and positioning accuracy. Highly consistent BRIRs needed to be acquired at different locations and a large amount of subjects was involved in the listening experiments. Hence, the use of individual HRTFs was not feasible at all. No individualization is applied, since active individualization might generate unexpected artifacts and affect the test results. All listening experiments conducted for this thesis compare to given external references and do not rely on an internal reference. Hence, using individual HRTFs or individualization of the dummy-head HRTFs might indeed enhance the overall listening experience but it is supposed to have only minor influences on the results of the performed listening tests.

The approach presented in this thesis directly permits using individual HRTFs in array-based binaural systems. Due to the specific wave order adaptation using spatial subsampling of the HRTFs individual HRTFs can be acquired and integrated effortlessly. For a reasonable array-based binaural system working at decomposition order of $N = 7$ and using a Gauss composite grid, we only need to acquire $M_{\mathrm{cg}} = 128$ (individual) HRTFs distributed over the entire sphere. For lower decomposition orders (e.g. $N = 5$) and more efficient composite grids (e.g. Lebedev), this number be reduced (e.g. $M_{\mathrm{cg}} = 50$) in theory, with considerable perceptual losses, cf. Section 5.

For concrete applications, each listener could use individual HRTFs in his personal rendering instance for listening to dynamic binaural sound recordings or productions.

Dynamic listening refers to head-tracked reproduction in at least three different degrees-of-freedom that are inherently covered by the modal description.

3.12.2 Torso reflections

The Neumann KU 100 that was used to acquire all binaural data in this work consists of a head only, in contrast to other models, such as the KEMAR Manikin (Burkhard and Sachs, 1975) or F.A.B.I.A.N. (Lindau and Weinzierl, 2006), that use a full torso. The influence of the torso was analyzed by Algazi et al. (2001a) and Guldenschuh et al. (2008), for example. The torso reflections are primarily evaluated by the auditory system for stabilizing the localization of sound sources in the median plane, i.e. for estimating the height of sound sources. For this work torso reflections are not considered.

3.12.3 Discrete Spatial Sampling

HRTF measurements are performed for discrete spatial sampling positions. The required number and the arrangement of different source directions to be acquired are important parameters.

For the modal adaptation of HRTFs to order-limited source systems, such as microphone arrays or higher-order Ambisonics (HOA) (Gerzon, 1985), (Daniel et al., 2003) decoders, spatial subsampling of the HRTF set was found to deliver best results, refer to Section 3.5.2. The spatial subsampling of the HRTFs in the space-frequency domain means modal low-pass filtering in the spherical wave spectrum domain.

The required number and position of sampling nodes are precisely determined by the composite grid that is used for performing the modal adaptation. The composite grid, in turn, is determined by the resolution of the source system that delivers the modal sound field description. The composite grid should be a quadrature that provides the same modal order as the underlying sound field description. The only remaining variable is the type of quadrature to be used as composite grid. From a technical or mathematical point of view, the differences between different composite grid types are low. In contrast, from a perceptual point of view the differences are indeed considerable, which is shown in Section 5. The equiangular Gauss quadrature has convenient and reliable properties in this context. Hence, we generally use Gauss quadratures as composite grids. Gauss quadratures provide a grid efficiency of $\eta_\mathrm{g} = 2$.

Thus, we need to capture exactly $2\,(N+1)^2$ HRTFs distributed over the entire sphere according to the Gauss quadrature specification in order to auralize a modal sound field description of order N.

3.12.4 Spherical Harmonic Interpolation of HRTFs

As stated before, a full HRTF $H^{\mathrm{l,r}}(\theta, \phi, \omega)$ set is described by a complex function of magnitude and phase on a virtual sphere S_h. Hence, the spatial Fourier transform is applicable. The result is a spherical harmonic expansion $\mathring{H}^{\mathrm{l,r}}_{nm}(\omega)$ of the underlying HRTF set. Measuring HRTFs is done by discrete spatial sampling. As a consequence, the discrete spatial Fourier transform from Eq. (3.3) applies de facto yielding

$$\mathring{H}^{\mathrm{l,r}}_{nm}(\omega) = \sum_{g_{\mathrm{sg}}=1}^{M_{\mathrm{sg}}} w_{g_{\mathrm{sg}}}\, H^{\mathrm{l,r}}(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}}, \omega)\, Y_n^m(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})^*. \tag{3.68}$$

The required spatial sampling nodes $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})$ and weights $w_{g_{\mathrm{sg}}}$ with $g \in [1, M_{\mathrm{sg}}]$ are determined by a suitable quadrature. The specific requirements are discussed below.

The spatial Fourier coefficients $\mathring{H}^{\mathrm{l,r}}_{nm}(\omega)$ contain sufficient information for performing continuous interpolation over the surface of the virtual sphere S_h, which is discussed in Section 2.15.

The inverse spatial Fourier transform, cf. Eq. (2.96) can be applied, yielding

$$\tilde{H}^{\mathrm{l,r}}(\theta, \phi, \omega) = \sum_{n=0}^{N \leq N_{\mathrm{sg}}} \sum_{m=-n}^{n} \mathring{H}^{\mathrm{l,r}}_{nm}(\omega)\, Y_n^m(\theta, \phi), \tag{3.69}$$

where $\tilde{H}^{\mathrm{l,r}}(\theta, \phi, \omega)$ represents a SH-interpolated HRTF. The maximum transform order N is determined by the order of the sampling quadrature N_{sg}.

This yields a continuous representation for any angle $(\theta, \phi) \in S$ from a discrete subset $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}}) \in S$, $g \in [1, M_{\mathrm{sg}}]$ by spherical harmonic interpolation. In other words, we obtain HRTFs for arbitrary angle (θ, ϕ) that can be, but does not have to be, contained in the original data set $(\theta_{g_{\mathrm{sg}}}, \phi_{g_{\mathrm{sg}}})$, $g \in [1, M_{\mathrm{sg}}]$.

To determine the required transform order N and the grid order N_{sg} we perform an approximation in the following. We span a virtual sphere S_h around the center of the head whose radius is delimited by the pinnae. As a reasonable choice we assume a sphere radius of $r_h = 9.5\,\mathrm{cm}$. Let us assume 20 kHz to be the highest temporal frequency to be resolved. If we insert this into (Rafaely, 2005)

$$N = \frac{\omega}{c}\, r_h, \tag{3.70}$$

we come to the conclusion that the HRTF set should be completely described using a maximum spherical harmonic order of $N = 35$ in theory. Figure 3.6 illustrates the modal intensity distribution of a measured HRTF set and indeed confirms our theoretical assumptions. As a consequence, the HRTF set needs to be sampled using a quadrature of $N_{\mathrm{sg}} \geq 35$, in order to obtain a suitable spherical harmonic expansion

and to perform spherical harmonic interpolation that is valid on the entire audible spectrum up to 20 kHz. A Lebedev quadrature of $N_{\mathrm{sg}} = 35$ with $\eta_{\mathrm{g}} = 1.3$ yields 1730 spatial sampling nodes or source positions, respectively.

Even though for this grid configurations the aliasing frequency is shifted to the outer limit of the considered temporal frequency range, we must still expect minor aliasing contributions over the entire spectrotemporal range due to the fact that natural sound fields are not order-limited. This is discussed in Section 3.8. However, the additive spatial noise from orders greater than N_{sg} is negligible in practice.

In order to keep a safety margin and to decrease spatial aliasing artifacts, a slightly higher grid order should be used in practice. Therefore, we use a Lebedev quadrature of order $N_{\mathrm{sg}} = 41$ with 2354 spatial sampling nodes in the following. Note that arbitrary grid order $N_{\mathrm{sg}} \geq 35$ is valid here. The data set for generating the subsequent content is described in Section 4.3.2.

The following figures give practical insight into spherical harmonic HRTF interpolation. First, we use different transform orders N to observe the development of an interpolated HRTF. Even using a HRTF set that is sampled at $N_{\mathrm{sg}} = 41$, the interpolation order N can be varied while fulfilling $N \leq N_{\mathrm{sg}}$, compare Eq. (3.69). Figure 3.60 shows an interpolated HRTF $\tilde{H}^1(\pi/2, \pi/4, \omega)$ at different interpolation orders $N = [5, 10, 20, 35]$ versus the original measured counterpart $H^1(\pi/2, \pi/4, \omega)$ from a different measurement session.

Indeed we observe that at $N = 35$ the interpolated HRTF comes very close to its original measured counterpart. For orders $N < 35$ specific impairments arise, like low-pass filter effects. This phenomenon was discussed in Section 3.5 in the context of a binaural system with limited modal resolution.

This example is a favorable one. The interpolation performance depends on the exact source position. Therefore, in Figure 3.61 additional examples with different source positions ($\theta = \pi/2, \phi = [0, \pi/4, \pi/2, 3\pi/2]$ interpolated at $N = 35$ are provided. The figure also comprises unfavorable cases.

We notice that deviations arise at very high frequencies and particularly at sharp notches in the magnitude response. The deviations reflect the impact of several secondary factors such as minor residual information in modal orders $N > 35$, spatial aliasing contributions, small motor positioning errors, temperature fluctuations during a measurement session, or temperature differences between the two measurement sessions, background noise, numerical errors, and so forth. Considering this large variety of influencing factors we can conclude that spherical harmonic interpolation of HRTFs works astonishing well in practice.

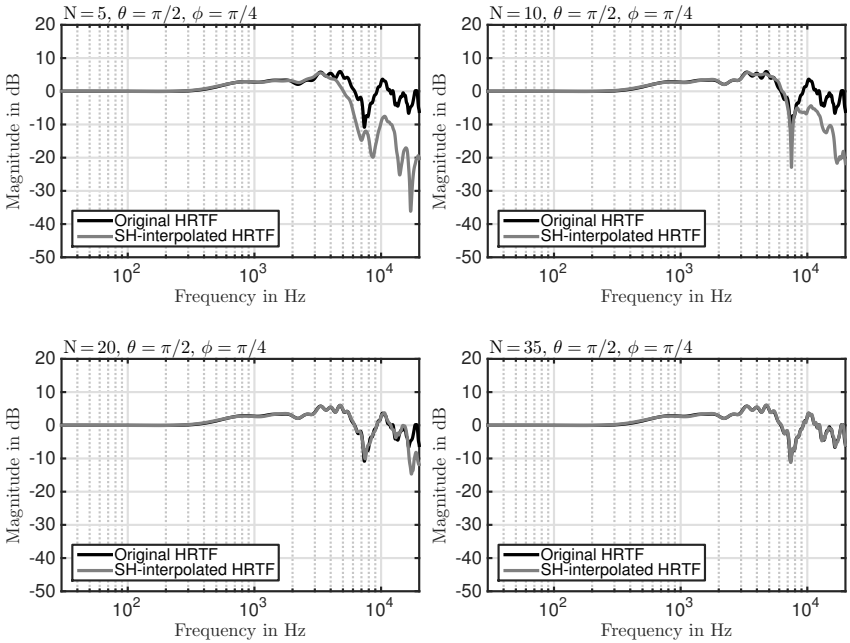


Figure 3.60 Original HRTF $H^{\mathrm{l}}(\pi/2, \pi/4, \omega)$ and interpolated HRTF $\tilde{H}^{\mathrm{l}}(\pi/2, \pi/4, \omega)$ at different interpolation orders.

Even though in a strict sense we only find that $\tilde{H}^{\mathrm{l,r}}(\theta, \phi, \omega)$ approximates $H^{\mathrm{l,r}}(\theta, \phi, \omega)$, the differences are negligible in practice and can be considered imperceivable. Therefore, throughout this thesis we assume $\tilde{H}^{\mathrm{l,r}}(\theta, \phi, \omega) = H^{\mathrm{l,r}}(\theta, \phi, \omega)$.

Spherical harmonic interpolation is a useful approach for deriving the required HRTF angles defined by the composite grid nodes $(\theta_{g_{\mathrm{cg}}}, \phi_{g_{\mathrm{cg}}})$ in Eq. (3.10) from a common high-resolution HRTF set that originally does not provide the required node positions. This operation is inherently contained in Eq. (3.11). Both equations adapt the HRTFs to a modal sound field description of limited modal order by performing subsampling of the HRTF set. This approach is applied for all simulations and listening experiments throughout this thesis. Respective routines are implemented in the SOFiA toolbox, refer to Section 4.2.

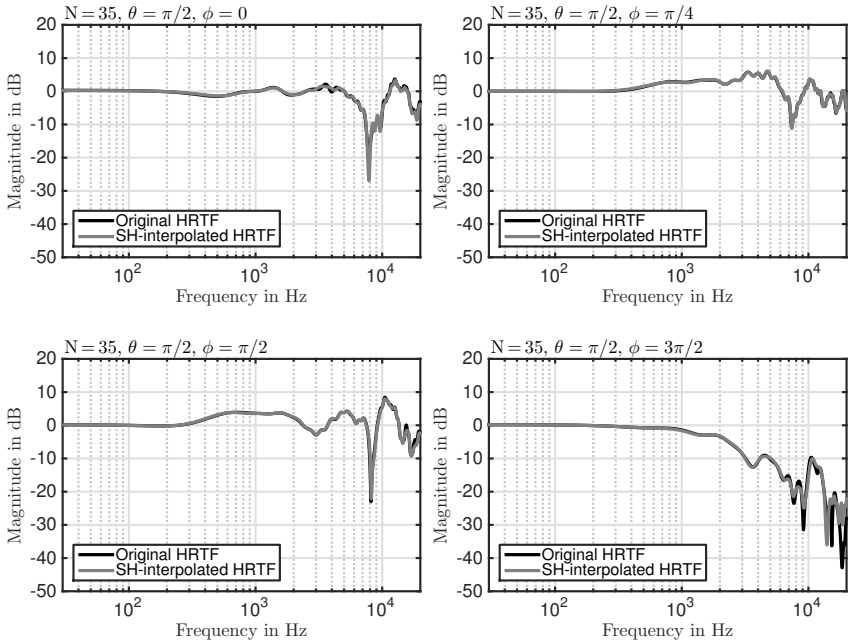


Figure 3.61 Original HRTFs $H^1(\theta, \phi, \omega)$ and interpolated HRTFs $\tilde{H}^1(\theta, \phi, \omega)$ at interpolation order $N = 35$ for different source directions.

More details on spherical harmonic expansion or interpolation of HRTFs are discussed e.g. by Evans et al. (1997), Nelson and Kahana (2001), Duraiswami et al. (2004), Zhang et al. (2009), Zotkin et al. (2009), Pollow (2010), or Zhang et al. (2012).

Range extrapolation of HRTFs can be performed using similar approaches. This not treated here but discussed by Duraiswami et al. (2004), Pollow (2010), or Spors et al. (2012a), for example.

3.13 Additional Factors of Influence

Major factors of influence, such as limited modal resolution of the sound field description, limiting the radial filter gain, discrete spatial sampling, and uncorrelated noise, were discussed so far. Those factors are found to be most crucial for properly assessing and describing the transmission properties of an array-based binaural system.

However, there are several additional factors that are not discussed at this point, since an appropriate in-depth discussion of all possible factors would go beyond the scope of this work. However, some of the additional factors are briefly mentioned. Several of them were already analyzed by different authors. Most of the other analyses focus on the influence of the respective factors on the PWD or beamformer only. As we see in Section 3.7 (noise analysis), the impact on the array working in composite mode might be considerably different from the impact on the isolated PWD. Thus, most of the analyses that are found in the literature may be enlightening for any kind of beamforming applications, but might not be conclusive in the context of array-based binaural auralization. In most cases they need to be repeated with regard of the composite processing in order to obtain meaningful results for this specific purpose.

For analyzing further factors of influence, the approach could be similar to the one presented in Section 3.7. The PWD signal, the composite signal, as well as the final binaural output signal, should be analyzed separately in order to deliver conclusive information regarding the impact on array-based binaural auralization.

3.13.1 Sources in the Near-field

The mathematical approach yielding a theoretically transparent behavior of the array-based binaural system that is discussed in Section 2.20 and followed throughout the entire thesis is only fully valid for sources in the far-field. Sources that are located close to the array yield certain impairments. The specific adaptation of the approach to a source in the near-field is not a problem. The radial filters can be adapted to resolve a near-field source properly.

However, there is a fundamental problem concerning near-field sources. In order to adapt the radial filters to a near-field source, the distance of the source must be known. Unfortunately, we usually do not know the positions of sources in the sound field. Furthermore, the radial filters need to be specifically focused on a dedicated distance, comparable to a camera lens that is focused an a specific focal plane. As a consequence, there is no general solution to generate radial filters that are capable of perfectly resolving sources at arbitrary distances. The radial filters used throughout this work focus the array to infinity, which appears to be the most general choice for processing unknown sound fields.

Near-field sources in the context of spherical microphone arrays are discussed e.g. in (Kennedy et al., 1996), (Abhayapala et al., 1999), (Fisher and Rafaely, 2008), (Abhayapala, 2008), (Fisher and Rafaely, 2009), or (Fisher and Rafaely, 2011).

3.13.2 Transducer Positioning Errors

The microphones might not be perfectly located at their assigned spatial positions in practice. Positioning errors of the array transducers yield orthogonality errors, as the spherical harmonic base functions are not sampled at the assigned positions and hence the acquired magnitude and phase information does not correspond to the aimed sampling nodes. The orthogonality errors are particularly critical at both very low and very high wavelengths compared to the array radius. Positioning errors are e.g. discussed and analyzed by Rafaely (2005).

3.13.3 Non-ideal Transducers and Interindividual Differences

Real microphones show a deviation from the ideal sensitivity, frequency response, and directivity. Moreover, the single array transducers show interindividual differences. The resulting errors are hard to quantify and to express in reasonable measures, since the variation possibilities are virtually infinite. Helwani et al. (2011) and Rettberg et al. (2012) propose using a multiple input – multiple output (MIMO) system theory approach for the calibration of microphone arrays in practice.

3.13.4 Time-variances

When using a scanning array with a single transducer instead of a real-time array with multiple microphones, amplifiers and ADCs, the interindividual differences largely vanish. Nevertheless, the single transducer still shows non-ideal properties. We strictly assume time-invariant properties of the medium and surroundings during the sequential acquisition of the complete array response. However, there are several factors that might not be time-invariant in practice, e.g. background noise, air movement, or varying characteristics of the speaker system due to increasing heat generation in the drivers. Another relevant factor is the air temperature, which due to natural environmental influences or air conditioners might have considerable fluctuations during the sequential acquisition of a complete array response. The influence of time-variances in the air temperature on scanning arrays is analyzed in (Bernschütz et al., 2011c).

3.13.5 Incomplete Sampling

A particular problem is incomplete or non-uniform sampling, where either single nodes or entire sections of the sampling grid are missing or the nodes are distributed very irregularly. Zotter (2009b) declares non-uniform and incomplete sampling as «ultimate challenge» in spatial sampling. Indeed, both are non-trivial problems demanding for complex theoretical or numerical approaches. Incomplete and non-uniform sampling

are discussed e.g. in (Pail et al., 2001) and (Du et al., 2003). More recent approaches to this topic based on sparse representation theory are discussed e.g. in (Rauhut and Ward, 2011) or (McEwen et al., 2013).

3.13.6 Non-ideal Sphere in the Measurement System

Unspecific deviations were observed in the measured array data sets, particularly for frequencies above 6 kHz. Spatiotemporal errors arise, such as minor ghost images in the space-frequency domain, as well as decreased output magnitudes of the binaural signal. These phenomena can unambiguously be ascribed to the non-ideal measurement setup. The data sets were captured using the VariSphear scanning array (cf. Section 4.1) equipped with a rigid sphere head. The setup is depicted in Figure 3.62.

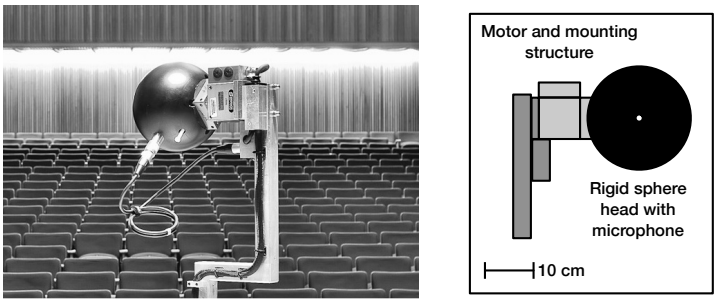


Figure 3.62 VariSphear scanning array with rigid-sphere head that provides a sphere diameter of $d = 17.5\,\mathrm{cm}$. The picture [Foto: P. Stade] (left) and the true-to-scale sketch (right) show the rigid sphere, the motor, and the mounting structure.

The motor sticks out of the sphere and the mounting structure is quite close to the measurement head. Both motor and mounting structure interact with the sound field, yielding non-ideal acoustical properties of the sphere. The motor and the mounting structure are not even fix obstacles in the sound field, but their position changes with the position of the respective target sampling node. As a consequence, the interaction is highly complex and can only be modeled using a boundary elements method (BEM) in a reasonable way. But even though the interaction is modeled properly, the theoretical approach needs to be modified accordingly. This was not within the scope of this work. Hence, neither the data nor the theoretical approach are modified to account for the latter. As a consequence, the non-ideal properties of the measurement setup yield certain errors during the processing, since ideal properties are assumed. The

errors increase towards high temporal frequencies, as the wavelength is short and the additional obstacles become an increasingly significant factor of influence.

A decreased output level of the binaural signal at high temporal frequencies was observed that is quite comparable to a high-shelf filter. In order to compensate for the resulting global coloration, a suitable compensation filter was designed that at least equalizes the diffuse field response of the array-based binaural signals. The filter magnitude is depicted in Figure 3.63. The filter was derived from comparing the circular diffuse field response of the original BRIR sets with the circular diffuse field response of the corresponding array based binaural room impulse response (ABRIR) sets and is consistently applied to all ABRIRs involved in the listening experiments presented in Section 5.7.

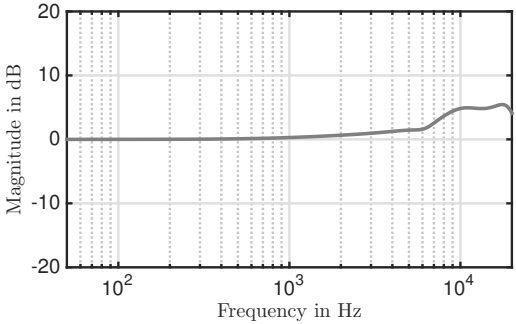


Figure 3.63 Spectral error compensation filter for a decomposition order of $N = 7$. The filter is applied to the array based binaural room impulse responses (ABRIRs) in Section 5.7 and aims to compensate for the resulting coloration due to the non-ideal sphere configuration of the VariSphear rigid sphere head.

This approach is the only feasible way to cope with the non-ideal properties of the specific measurement system in practice. For arrays that are closer to an ideal sphere, the spectral compensation filter is not required. We must keep in mind that the listening experiments based on measured data (Section 5.7) are negatively influenced by the errors introduced by this specific measurement system, even though the overall diffuse field response is compensated. Therefore, the ratings can generally be expected to be slightly higher for arrays with more ideal properties, i.e. providing a sphere without motors sticking out or comparably large mounting structures being close to the measurement head.

4 Technology and Resources

This chapter describes and discusses the technology and resources that were used for the experiments and simulations. The field of spherical microphone arrays for auralization –and spherical acoustics in general– can be considered to be a young discipline in science. The relevance for commercial applications is still comparably low at the current state of research and technology. Thus, hardly any suitable hardware, software or measured data is available. As a consequence, considerable effort needed to be spent for the design and construction of specific hardware, the design and implementation of signal processing algorithms, acquiring measured data, and setting up a suitable environment for listening experiments. All software or data resources that were built or acquired are made available to the scientific community.

4.1 VariSphear Scanning Array

A variable spherical scanning microphone array measurement system, called VariSphear[1], was developed and constructed (Bernschütz et al., 2010). Scanning microphone arrays are discussed in Section 3.2.1. The VariSphear systems consist of specifically designed hard- and software for the acquisition of array impulse responses or array room impulse responses. After completing the prototype, several additional VariSphear measurement systems were built for universities and industrial companies. The VariSphear array was e.g. used for the extensive measurement sessions in the SEACEN project (Weinzierl et al., 2012).

4.1.1 Hardware

The hardware of the array consists of a stable ground plane with leveling feet, carrying a robot arm structure that meets a best possible compromise between constructional stability and acoustical transparency. The construction is made of aluminum. The array hardware is depicted in Figure 4.1. The robot arm provides two motorized degrees of freedom (azimuth ϕ and elevation θ), as well as two manual degrees of freedom (sphere radius r_0 and measurement height h_0). Two Schunk/Amtec Robotics PR 70 motors with internal position sensors and position references, as well as magnetic breaks are used for realizing the automatized degrees of freedom. The construction provides very

[1] VariSphear: Variable Spherical Ear.

good repetition accuracy with angle errors $|\Delta\Omega| < 0.01°$. The array can be changed from open sphere configuration with variable radius to rigid sphere configurations with a fixed radius. The VariSphear system provides a Hygrosens TSIC-LABKIT/TSIC 306 temperature sensor for tracking the air temperature during the measurements. A remotable TOPCON LEM 30 laser distance sensor can be mounted to the robot arm for capturing distances and angles for simple true-to-scale CAD modeling. All motors and sensors are tied to a MOXA Nport 5410 industrial device server, providing Ethernet connectivity to the external world for controlling the entire measurement system. Nearly all parts of the system, except the electronics, were custom built from raw material in the mechanical workshop of the Institute of Communications Systems at Cologne University of Applied Sciences.

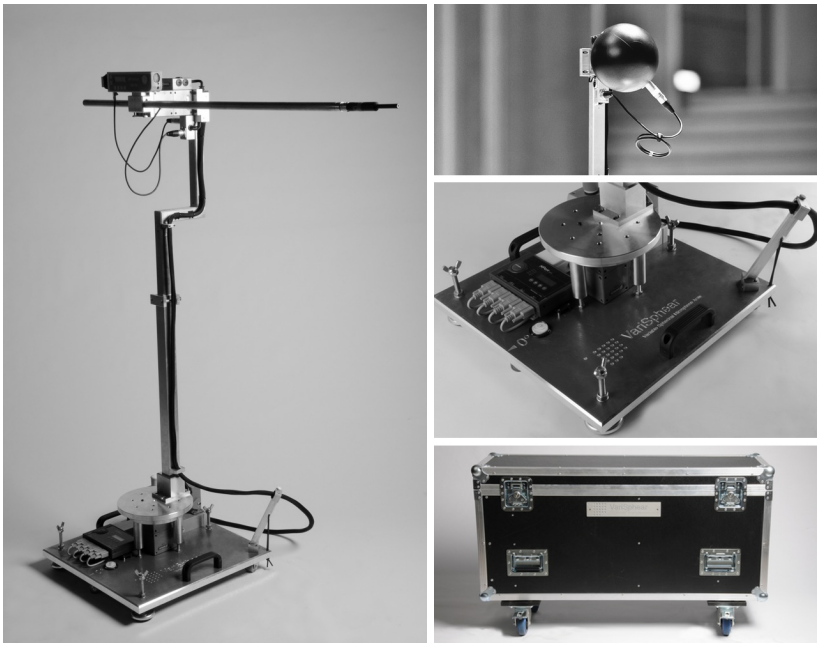


Figure 4.1 Images of the VariSphear hardware including the rigid sphere head and the flightcase for transport [Pictures: S. Moritz, except right, top: P. Stade].

Several additional heads and mounts for using the motion base in different applications were built. The custom mount for capturing spherical head-related transfer functions (HRTFs) or binaural room impulse responses (BRIRs) is depicted in Figure 4.5 on page 168. For taking spherical panorama pictures, a pano head was developed for mounting remotable reflex cameras. The pano head can be seen in Figure 4.19 on page 181. The laser head for acquiring CAD data is depicted in Figure 4.20 on page 181.

4.1.2 Software

For controlling the hardware and performing automatized impulse response measurements, a proprietary software was developed and implemented. The software is written under MATLAB® and offers a comprehensive guided user interface (GUI), which is depicted in Figure 4.2.

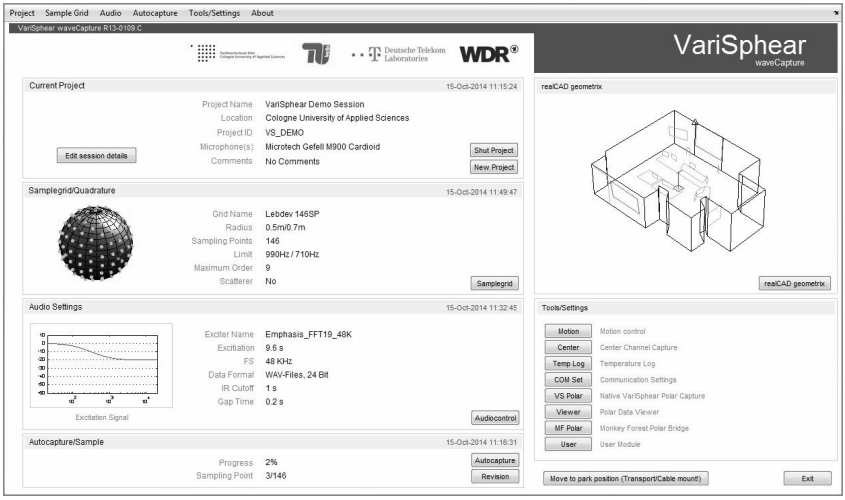


Figure 4.2 VariSphear software; screenshot of the main page, giving an overview of the current session settings and status.

The reader is referred to the VariSphear manual (Bernschütz, 2013b) for a detailed description of the software. The following listing gives a short overview of the most important features provided by the VariSphear software:

- Project/session management
- Automatized capturing of array impulse responses including meta data

- Automatic measurement error detection and correction (e.g. audio dropouts, background noise events, motion/hardware errors)
- Data revision module for verification of the measurements and manual replacement of single positions
- Impulse response capturing core with port audio binding, based on sine sweep excitation and integrated deconvolution according to Müller (1999).
- Predefined or user defined sampling grids/quadratures
- Proprietary CAD module (realCAD geometrix) for capturing, drawing, and editing true-to-scale models of the venue using the VariSphear laser head
- Manual motion/hardware control
- Module for capturing center impulse responses
- Polar data capturing and viewer, Monkey Forest polar bridge

In addition to the compiled GUI based software package, dedicated open MATLAB® functions are provided for controlling the VariSphear hardware and building proprietary scripts.

4.2 SOFiA Sound Field Analysis Toolbox

An open source sound field analysis toolbox called SOFiA[2] for MATLAB® was developed, implemented, and published (Bernschütz et al., 2011a). SOFiA provides an ample function set for simulating spherical microphone arrays, processing measured array data, analyzing and visualizing array responses, or performing binaural auralization based on analytic or measured array data sets. Since its first release and publication the toolbox was extended. The current version of the toolbox facilitates to transfer large parts of the theory discussed in Section 2 and Section 3 to practice in terms of simulation, visualization, and auralization.

SOFiA is implemented in MATLAB®. Certain operations with extensive computational demands are implemented as externals in C++ and integrated to MATLAB® via MEX[3] binding. This approach achieves improved computational performance while staying in the comfortable MATLAB® environment.

The toolbox was verified in extensive experiments and simulations. The software is open source under a GNU GPL v3 license and is used by other researchers, such as Spors et al. (2012b), Rettberg and Spors (2013), Schultz and Spors (2013), or Muhammad et al. (2014).

All simulation, visualization, or auralization conducted in this thesis is processed using SOFiA. A detailed discussion of the toolbox goes beyond the scope of this section. Hence, the description is limited to a signal flow diagram that is depicted in Figure 4.3. For more detailed information on SOFiA, the reader is referred to (Bernschütz et al., 2011a) and to the SOFiA code hosting page that can be found in the publication.

[2] SOFiA: sound field analysis
[3] MEX: MATLAB external

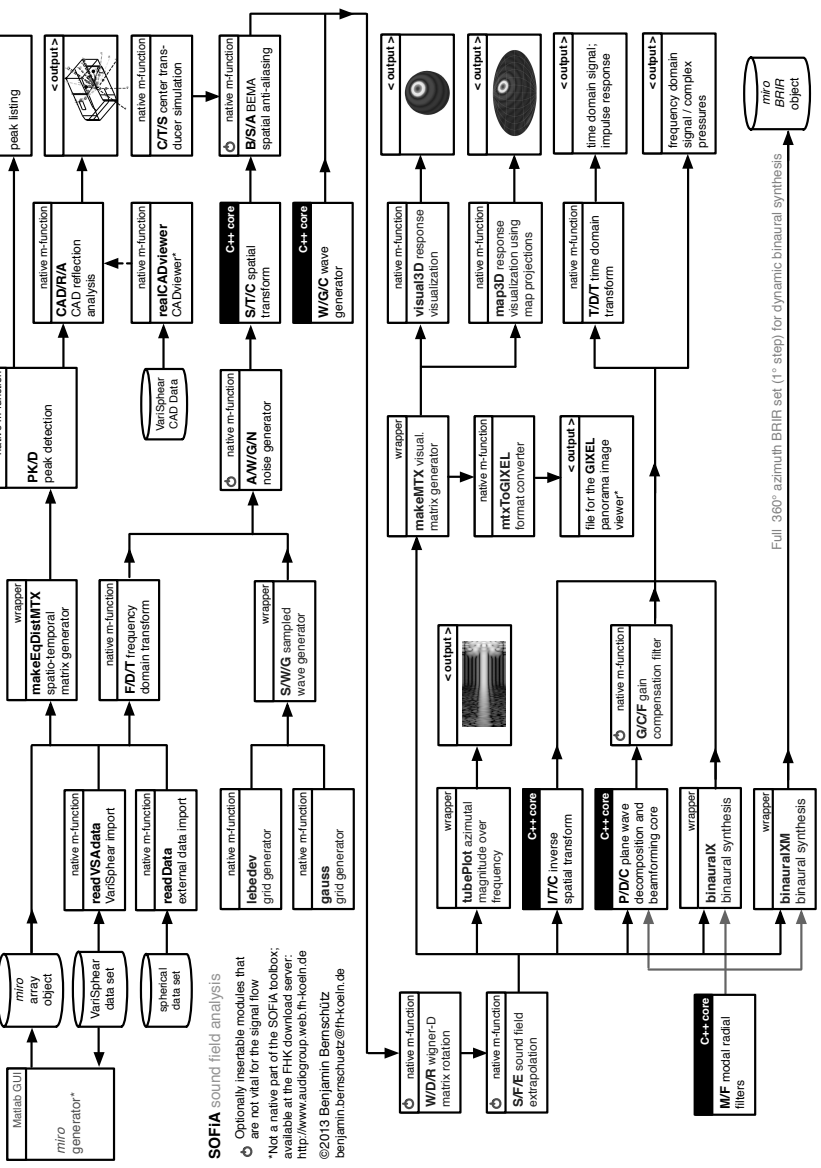


Figure 4.3 SOFiA R13-0306 signal flow diagram.

4.3 Measured Data

Several measurements of head-related impulse responses (HRIRs), binaural room impulse responses (BRIRs), and array room impulse responses (ARIRs) were performed. The data used for simulations and listening experiments in this thesis was acquired from scratch under strictest criteria.

Suitability of the acquired data was ascertained by precedent analytic simulations and dedicated strategic planning, in order to determine suitable setups and parameters for the measurement sessions.

For ascertaining validity of the data, the entire impulse response measurement chain was verified in several test procedures. During the measurement sessions, every individual impulse response in the data sets was reviewed and checked for abnormalities using specifically designed algorithms comparing the similarity to neighboring impulse responses and analyzing the signal-to-noise ratios (SNRs). In addition, a visual inspection of the impulse responses and the corresponding spectra was performed. Very few impulse responses turned out to be faulty due to clicks or dropouts in the audio chain or environmental noise during the data acquisition. The outliers could be clearly detected and the affected impulse responses were recaptured. The integrity of the data concerning completeness, indexing and meta data assignment (e.g. assignment of spatial directions) was checked by analyzing the deviations for neighboring impulse responses paired with a visual inspection of polar or balloon-plots showing magnitude and phase of the spatial data set.

Consistency of all data sets was established by using the identical setup in every detail for all locations including the anechoic chamber. The sources and source positions were intentionally varied at the different locations, but identical sources and positions were used for a common subset of receivers. The adjustment and positioning of the receiver pivot point was performed with great accuracy using cross-grid lasers and observing the phase responses using a real time analyzer. The spatial origin of the array was equal to the predefined center of the dummy head and the position of the captured monophonic impulse response. Furthermore, the measurements of binaural room impulse responses (BRIRs) in the rooms and head-related impulse responses (HRIRs) in the anechoic chamber are consistent concerning the pivot point and the physical structure of the apparatus, by using identical equipment including the rotation mounts. The distances of the different source positions to the receiver within a location were kept constant with great accuracy. As a consequence, the different data sets for a single source type can be overlayed with nearly perfectly matching phase responses of the direct sound leading to constructive summation on the entire audio spectrum without comb filter effects. The constancy of extrinsic physical conditions, such as air temperature or

humidity, are not controllable for different locations. Both were tracked and recorded during the measurement sessions. The background noise was analyzed during the impulse response measurements. Single noise events above the constant background noise were detected and the respective measurements were repeated automatically.

The aim was to acquire data that is suitable for auralization at the level of broadcast studio production. This leads to specific demands for the achieved frequency bandwidth, linearity of the frequency response, and the signal-to-noise ratio (SNR). More effort needed to be spent on the data acquisition than would be sufficient for just answering most of the arising scientific questions. The frequency bandwidth was set to cover the range of approximately 40 Hz to 18 kHz. This leads to ambitious demands concerning the involved sources and the density of the measurement grids. All sources were tuned to provide a flat frequency response on axis. For studio monitors, a very strict target of ± 1 dB of maximum deviation on the entire bandwidth was allowed. For the portable sources, the maximum allowed deviation was scaled down to ± 3 dB. The impulse response to noise ratio was set to be greater than 80 dB. In some locations, SNR values up to 115 dB could be achieved for the full measurement bandwidth. The data sets were captured at a temporal sampling rate of 48 kHz and a word length of 24 bits. As no suitable omni-directional source with sufficient bandwidth and output power is available on the market, a new source called Sonic Ball was developed from scratch, cf. Section 4.3.3.2. In order to obtain suitable HRTFs including the critical low-frequency range, an algorithm for the low-frequency extension of HRTFs was developed, cf. Section 4.3.2.2.1. All in all, great effort was spent in order to acquire data sets with good technical audio transmission properties. As a consequence, even critical expert listeners (Category A, cf. Section 5.3) from the field of radio production were pleased by the auralization during the listening experiments.

4.3.1 MIRO Data Format

In order to provide appropriate and structured storage, localization, signal processing, and management of the more than 50.000 impulse responses (IRs) that were acquired, a proprietary object-oriented data type called MIRO (Measured Impulse Response Object) for MATLAB® was developed. The structure of the MIRO data type is depicted in Figure 4.4.

MIRO was specifically designed for storing head-related impulse response (HRIR), binaural room impulse response (BRIR), or array based binaural room impulse response (ABRIR) sets. In addition to the measured data, approximately 50 different properties including meta data and adjustable processing parameters are provided in each instance. The MIRO class offers an internal pseudo real-time signal processing

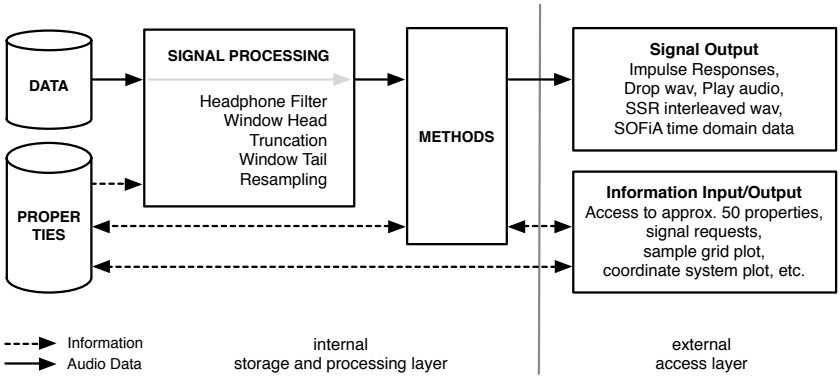


Figure 4.4 MIRO data type overview.

core. The processing is non-destructive and applied to the output signal just at the moment an IR is requested from the storage via its dedicated getter method. The processing core includes the application of headphone filters, adjustable truncation and windowing, as well as temporal resampling of the requested IR. The processing parameters are set via setter methods and written to the properties of the instance. Any other properties (also including the processing parameters) can directly be read or be written to the properties of the instance without using specific setter and getter methods. Even if this approach is slightly more unsafe than providing getter and setter methods, the usage is considerably more comfortable. Safety is not that important in the given context, since the objects are conceived as data storage containers for read-only data and thus an accidentally misconfigured or destroyed object can just be reloaded without loosing important information. The MIRO class offers several other methods that are listed in the following:

- Locate the closest matching impulse response to a given angle pair,
- Plot the underlying reference coordinate system,
- Plot or return the sample grid of the stored data in the specific instance,
- Pre-listen an adressed IR convolved with a passed audio stimulus,
- Drop a common mono or stereo wave file containing the addressed IR,
- Export interleaved HRIR or BRIR wave files for the Sound Scape Renderer (SSR) (Geier et al., 2008),
- Export SOFiA time domain data sets for seamless processing in the spherical harmonic domain using the SOFiA toolbox, cf. Section 4.2.

The MIRO class definition for MATLAB® is included in the published impulse response data sets introduced in Section 4.3.2 and Section 4.3.3. The MIRO data type, for example, was also adopted by the BBC spatial audio research group by Melchior et al. (2014) and Satongar et al. (2014).

A comparable data format named SOFA is under development (Majdak et al., 2013). It was initiated by the ABBA research group (Blauert et al., 2010). In the meantime, the SOFA format has considerable afflux from various international research groups and was recently released as standard (AES, 2015).

At the time of acquiring the data sets for the present work, the SOFA format was still unknown and rather elementary. Hence, the proprietary MIRO format was developed in parallel. In addition, the SOFA format does not offer a signal processing core or export methods to the SSR or to SOFiA in the current version. These are quite convenient features of the MIRO data type and justify the coexistence of SOFA and MIRO at the present time. Nevertheless, SOFA is most probably becoming the prevailing spatial audio data format and several additional features are expected. In order to make the data acquired in this thesis accessible to a broad community, the HRIR data sets from Section 4.3.2 were additionally made available in the SOFA format by the SOFA work group.

4.3.2 Far-field HRTF Measurements

Spherical far field HRTF sets with optimized properties for applications in the field of spherical acoustics were measured (Bernschütz, 2013a).

4.3.2.1 Environment and Setup

The measurements took place in the anechoic chamber at Cologne University of Applied Sciences. A *Neumann KU 100* dummy head was mounted on the VariSphear motion system, cf. Section 4.1. The custom-built rotation mount for the dummy head required a transform of the usual VariSphear coordinate system. The addressed virtual source positions were mapped to the motor steering angles. The coordinate transform was implemented as plug-in for the VariSphear software; hence all features, such as automated motion control, impulse response capturing, and error detection, were available for the HRTF measurements. An emphasized sweep with $+20\,\mathrm{dB}$ low-shelf at $100\,\mathrm{Hz}$ of 2^{19} samples was used for excitation. A Genelec 8260A speaker system was playing back the measurement signals at a distance of approximately $3.5\,\mathrm{m}$ to the center of the head. This distance is considered as far field referring to the speaker's dimensions. The speaker was tuned to a flat frequency response of $\pm 1\,\mathrm{dB}$ on axis. A

RME Fireface UCX was used as audio interface, including the built-in analog-digital converters, digital-analog converters and microphone pre-amplifiers.

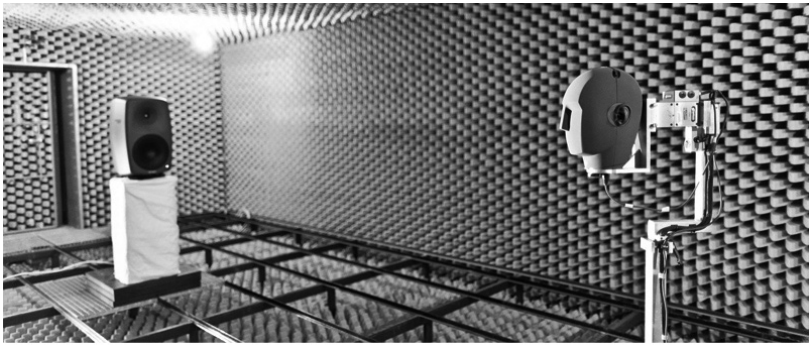


Figure 4.5 Setup for the HRTF measurements in the anechoic chamber [Picture: P. Stade].

The head positioning and angle adjustment were conducted accurately using cross-grid lasers and real-time phase analysis of the two ear signals with a real time audio analyzer. Different quadrature grids were captured:

- Circular azimutal grid with 1° stepsize (mic stand),
- Circular azimutal grid with 1° stepsize (full rotation mount),
- Spherical equidistant Lebedev quadrature with 2354 nodes,
- Spherical equidistant Lebedev quadrature with 2703 nodes, and
- Spherical equiangular 2° Gauss quadrature with 16020 nodes.

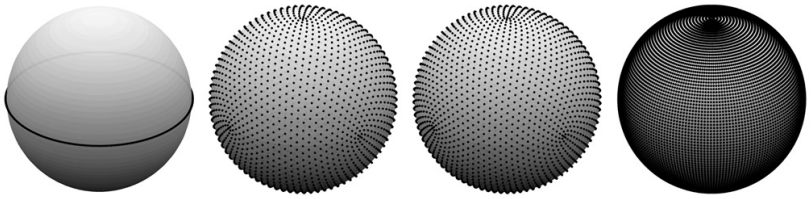


Figure 4.6 Spatial sampling configurations: Circular 1°, Lebedev 2354, Lebedev 2702 and Gauss-Legendre 2° (Bernschütz, 2013a).

The three spherical grids enable a stable transform of the respective HRTF sets to the spherical wave spectrum domain for frequencies up to at least 20 kHz with low spatial

alias contributions. Climate conditions in the anechoic chamber were tracked with temperature and humidity sensors. In order to detect time-variances in the measurement chain (e.g. due to driver heating) a fix-mounted measurement microphone on a separate channel was used. The temperature variations within the single measurement runs were below $\pm 0.5\,\mathrm{K}$ and no remarkable time-variance of the measurement chain was observed when comparing the progressive responses of the control channel.

4.3.2.2 Post Processing

In order to obtain a suitable HRTF data set with good audio-transmission properties referring to magnitude and phase, some post processing operations needed to be applied, which are briefly outlined in the following and described in (Bernschütz, 2013a) in greater detail. Besides the magnitude and phase properties of the HRTF that are inherent to binaural techniques (Møller, 1992), (Blauert, 1997), additional unwanted parasitic magnitude and phase changes arise owing to the measurement setup. If not considered and removed, the respective changes of the complex transfer function propagate through the entire binaural playback chain. A common problem is measuring the low frequency range of HRTFs. First of all, most anechoic chambers typically have a lower boundary frequency due to limited dimensions and length of the absorption wedges (Beranek and Sleeper Jnr, 1946; ISO, 2003) that is within the audible range. Below this frequency the chamber cannot hold the anechoic properties. As a consequence, room modes and reflections arise. In the present case, the anechoic chamber at Cologne University of Applied Sciences has a lower boundary frequency of approximately 200 Hz. Even if the very low frequency range does hardly influence the binaural hearing and localization (Blauert, 1997), (Møller, 1992), the reflections and room modes have substantial influences on the frequency response and the group delay properties of the measured HRTFs. Additionally, most speaker systems show ripple in the frequency response, high-pass characteristics and a considerable surge of group delay towards lower frequencies (Goertz, 2008), (D'Appolito, 1999), (Müller, 1999). High-pass characteristics of the HRTFs lead to a sound reproduction lacking low frequencies. Other changes or ripple in the magnitude response exceeding certain limits bring along audible coloration (Fastl and Zwicker, 2007, pp 175–202). Group delay distortions can cause audible impairment of the audio signal (Blauert and Laws, 1978) and lead to an unnecessary expansion of the time domain signal, i.e. the HRIR. This directly impacts the required computational power for the HRIR convolution. All those distortions should be considered and their effects removed in the post processing stage.

4.3.2.2.1 Adaptive Low Frequency Extension (ALFE) In order to avoid the severe problems due to measurement uncertainties at low frequencies, the low frequency components of the measured HRTFs were replaced by synthesized components. The procedure is similar to the approach presented by Xie (2009). A replacement is feasible, as the influence of the head on the sound field is negligible for low frequencies according to Rayleigh scattering for spheres of a dimension that is small compared to the wavelength (Bowman et al., 1970). The pinna filters also do not have a substantial effect on the low frequency range (Blauert, 1997). A simulation of a plane wave impact at different frequencies on a rigid sphere as a simplified head model was conducted in order to assess the actual scattering influences. Graphical results of the simulations are depicted in Figure 4.7, indeed indicating negligible scattering effects below approximately 200 Hz.

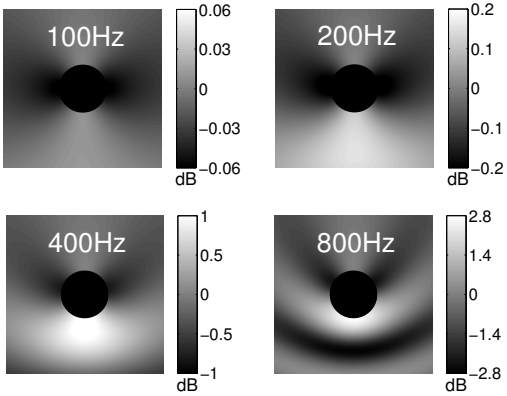


Figure 4.7 Simulated scattering effects for a monochromatic plane wave impact from south at different frequencies to a rigid sphere of a diameter d=17.5 cm that serves as simplified dummy head. The plots depict the resulting pressure magnitudes around the sphere. The pressure variations due to physical scattering effects are smaller than ±0.2 dB below 200 Hz, which can be neglected in practice.

An algorithm for dynamic replacement of the low frequency HRTF portion, called adaptive low frequency extension (ALFE), was developed. The block diagram of the ALFE algorithm is depicted in Figure 4.8.

The ALFE algorithm is independently applied to every single HRIR and HRIR channel. It first splits off a high frequency path of the raw HRIR, using a Linkwitz-Riley 24 dB/oct high-pass filter at the target crossover frequency of 200 Hz. The second path

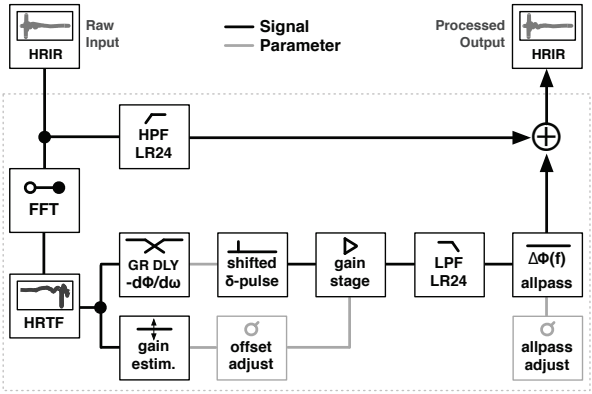


Figure 4.8 Block diagram of the ALFE algorithm from (Bernschütz, 2013a).

is transformed to the frequency domain and the group delay at the crossover point is determined.

A shifted dirac pulse is generated according to the group delay at the crossover point. The mean gain for the low frequency band is estimated by analyzing carefully selected frequency bins in the raw HRTF signal that do not match dedicated room modes. The dirac pulse is scaled in gain accordingly. At this point, a flat full-spectrum impulse with matching group delay and gain properties for the low frequency range is generated. In order to couple in this synthetic signal to the HRIR, a corresponding Linkwitz-Riley 24 dB/oct low-pass filter is applied. In a last step, all-pass filters are inserted in order to match the phase responses between the original high-pass filtered HRIR signal and the synthesized low-pass filtered ALFE signal around the crossover frequency. The phase-matching at the crossover frequency is necessary in order to obtain a constructive summation of both signals. The single output signals from the high-passed HRTF path, the low-passed ALFE path, and the sum signal are depicted in Figure 4.9 (top). The corresponding phase responses of the two signals, as well as the phase slope matching in the range of the crossover frequency can be observed in Figure 4.9 (bottom). Apart from the crossover frequency, the phase responses start to diverge, but the damping of the 24 dB/oct crossover provides sufficient attenuation. The output signal of the algorithm is a single HRIR channel with improved properties concerning the magnitude and phase responses compared to the raw HRIR channel at the input. The frequency response at low frequencies is flat and the group delay of the HRIR is reduced below the crossover frequency, which is illustrated in Figure 4.11. The high-pass characteristics

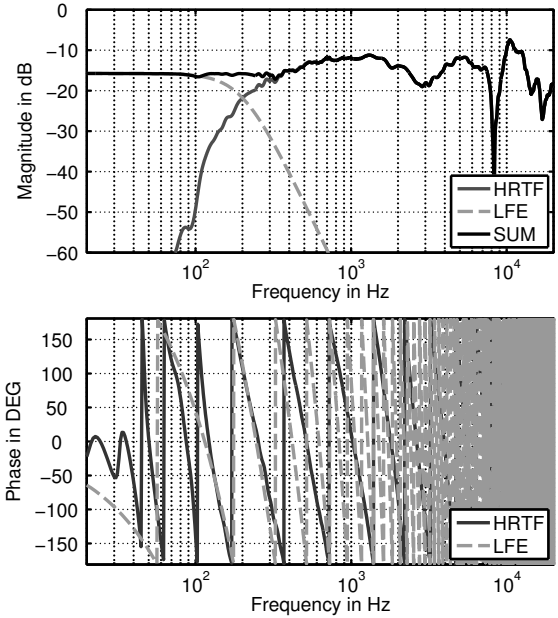


Figure 4.9 Exemplary magnitude and phase responses for the signal paths in the ALFE algorithm with a crossover frequency of 200 Hz.

is removed from the HRIR and thus a binaural system using this HRIR would be able to transmit low frequency content without restrictions.

4.3.2.2.2 Magnitude and Phase Compensation After replacing the low frequency range of the HRTFs, compensation filters for the magnitude and phase responses were generated. For that purpose, a transfer function was captured using a Microtech Gefell M296S measurement microphone at the pivot point instead of the dummy head. Appropriately smoothed inversion of the magnitude and phase response yields a suitable complex-phase FIR filter (Müller, 1999) that was used for compensating the ripple of the speaker's frequency response and to remove unnecessary group delay distortion in the transmission path. The complex-phase compensation filter is a non-causal filter, since it needs to remove lagging group delay portions in the transmission path. In combination with the measured HRTFs, non-causal parts clear away and only an overall frequency-independent latency remains. This latency can simply be cut off by removing the leading zeros. At that point, the time domain response (HRIR) is contracted

to a much more compact response, without affecting the inherent group delay variations that are necessary for binaural hearing. Using this approach, possible ripple in the magnitude response or group delay distortion of the transmission path are removed with a common compensation filter. Only the inverted properties (frequency and phase response) of the measurement microphone remain as an artifact in the transmission path, which usually can be neglected in practice. Hence, the measurement microphone should be as neutral as possible concerning its frequency and phase response for this compensation approach. The frequency and phase response of the transducers in the dummy head remain in the overall signal chain, as these are not involved in the compensation process. These are compensated during the headphone equalization that is introduced in Section 4.3.4.

The magnitude compensation is illustrated in Figure 4.10. The plot shows the magnitude responses of different paths of the ALFE algorithm from Section 4.3.2.2.1. The sum of both paths is the resulting HRTF without compensation filter. Some minor ripple in the magnitude response can be observed (black curve), that can be ascribed to the loudspeaker system. Since the loudspeaker system was already tuned to a very flat response with deviations smaller than ± 1 dB during the measurement session, the ripple is comparatively small and could be neglected in practice. But since the complex-phase compensation filter is applied for improving the group delay properties anyways, the remaining ripple is removed in the same step. The light-gray trace shows the resulting compensated magnitude response, where the loudspeaker's magnitude response is removed.

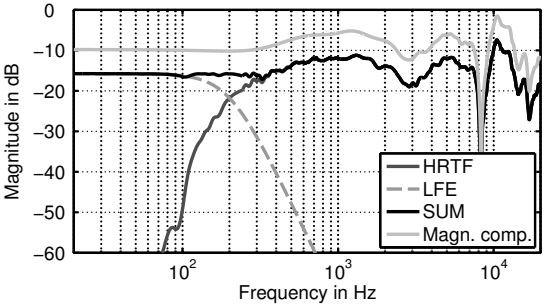


Figure 4.10 Compensation of the magnitude response.

The influence of the more essential phase response compensation on the group delay and the time domain are illustrated in Figure 4.11 and Figure 4.12. Figure 4.11 shows the group delay responses at three different stages of the post processing. The continuous black curve represents the raw measured HRTF. A tendential surge of group delay

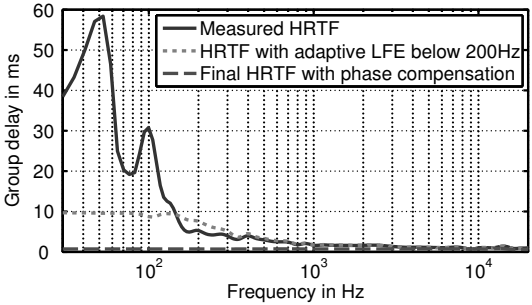


Figure 4.11 Group delay at different stages of the post processing.

with two additional peaks at 50 Hz and 100 Hz can be observed at low frequencies. The surge of group delay can be mainly ascribed to the speaker system and the two peaks are due to resonances in the anechoic chamber. The dotted gray curve represents the HRTF after applying the ALFE algorithm. The group delay is limited to maximum values of 10 ms on the entire frequency range. A slight increase of group delay can be observed around the ALFE crossover frequency of 200 Hz, which is a consequence of the Linwitz-Riley 24 dB/oct crossovers filters. The dashed black trace represents the final HRTF with applied complex-phase compensation filters.

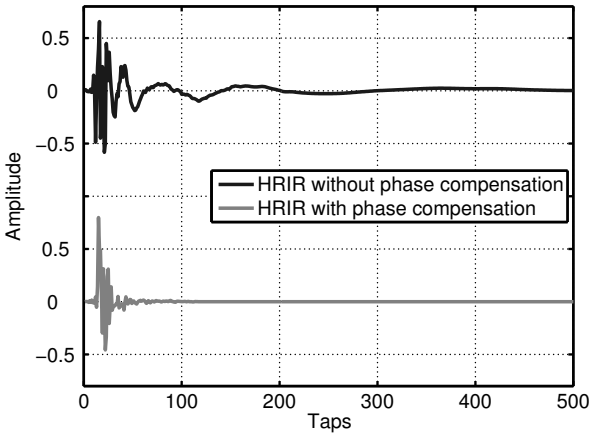


Figure 4.12 Time domain response without and with phase compensation.

Any surge of group delay that is not inherent to the binaural transfer function itself is removed at this point. The true advantage of the phase compensation filters achieving

a flattened group delay response of the HRTF can be observed in the time domain, illustrated in Figure 4.12. The black trace (top) shows a HRIR channel in the time domain with applied ALFE algorithm. According to the maximum group delay of 10 ms, the impulse response covers around 480 filters taps at a temporal sampling rate of 48 kHz, whereas the measured response without ALFE algorithm would even cover considerably more taps. The low frequencies are mainly coded in the tail section. Cutting off the impulse response at e.g. 128 taps would entail severe loss of low frequencies. Hence, the response should not be cut off at any value lower than 480 taps in order to maintain a reasonable frequency response at low frequencies. The gray trace (bottom) shows the identical HRIR with applied phase compensation filters. The response is clearly tightened. Cutting off this response at e.g. 128 taps, would not have any consequence in the frequency domain in contrast, since the response is entirely decayed. Hence, with phase compensation, the HRIRs can be much shorter. This saves memory and computational power during the convolution in a binaural system, while maintaining identical spectral properties and minimized group delay distortions.

4.3.3 WDR Spatial Audio Impulse Response Compilation

An extensive impulse response compilation was captured at the Westdeutscher Rundfunk (WDR) broadcast studios for this thesis, an affiliated master thesis (Stade, 2013), and an affiliated bachelor thesis (Rühl, 2012).

Figure 4.13 Poster for presenting the WDR spatial audio impulse response compilation at the 27th VDT International Convention in Cologne/Germany (Stade et al., 2012) [Picture: P. Stade].

Impulse response sets in several rooms using different receivers and sources were recorded. Additionally, CAD Models of the rooms and panoramic photographs were captured. The compilation is called WDR spatial audio impulse response compilation and published in (Stade et al., 2012). The final compilation involves more than 35.000 room impulse responses. The structure and procedure of the measurements is comparable to the extensive measurement sessions that were conducted in the SEACEN project (Weinzierl et al., 2012). In the following, some elementary information on the compilation, as well as some pictured impression are presented. For a more detailed description, the reader is referred to (Stade et al., 2012).

4.3.3.1 Rooms

Two studio control rooms are included in the compilation, i.e. control room 1 – CR1 (music production) with a room volume of $93\,\mathrm{m}^2$ and an average reverberation time

60 (RT60) of 0.23 s, as well as control room 7 – CR7 (radio drama production) with a room volume of 168 m^2 and an average RT60 of 0.25 s, cf. Figure 4.14. Two broadcast studios are included, i.e. the small broadcast studio - SBS with a room volume of 1250 m^2 and an average RT60 of 0.9 s, as well as the large broadcast studio with a room volume of 6500 m^2 and an average RT60 of 1.7 s, cf. Figure 4.15.

Figure 4.14 Control rooms: a.) Control room 1 – CR1 [Picture: WDR/Hagmayer], b.) Control room 7 – CR7 [Picture: WDR/Maurer].

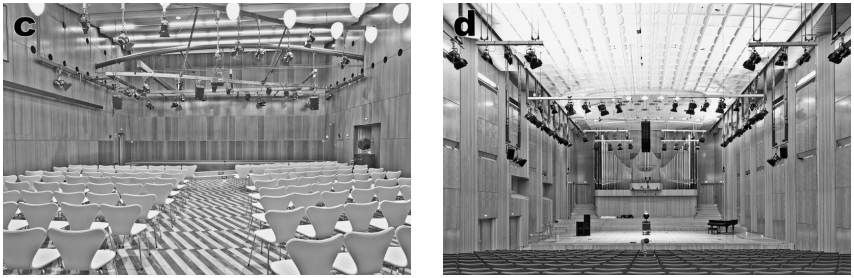


Figure 4.15 Broadcast studios: c.) Small broadcast studio - SBS and d.) Large broadcast studio - LBS [Pictures: P. Stade].

4.3.3.2 Sources

Different sound sources were used for exciting the rooms. In the control rooms, the available studio main monitor systems, Genelec 8260A or B&W 803D were used for that purpose, cf. Figure 4.16.

In the broadcast studios, portable sound systems with two different source characteristics were used for excitation. The first sound source was a PA stack from AD-Systems,

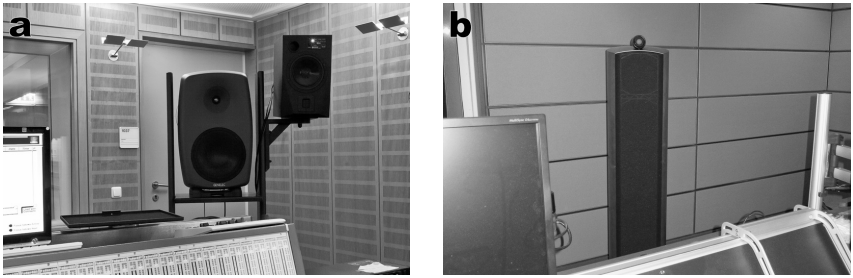


Figure 4.16 Sound sources in the control rooms: a.) Genelec 8260A in CR1 and b.) B&W 803D in CR7 [Pictures: P. Stade].

consisting of a horn-loaded 2x12" mid/high unit called Stium and 3x15" Flex-Series subwoofers, which is depicted in Figure 4.17 (c). The Stium speaker system has a

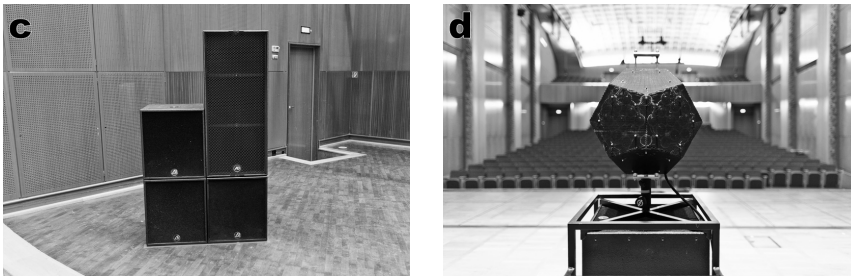


Figure 4.17 Sound sources in the broadcast studios: c.) AD Systems PA stack (PA) and d.) Sonic Ball omni-directional source (SB) [Pictures: P. Stade].

nominal pattern of 75° H x 50° V. The stack was driven by Camco Vortex amplifiers and controlled by an XTA system controller. It was aligned in the anechoic chamber in advance and the overall equalization was only sparsely adapted to the respective rooms, both using the Monkey Forest measurement system from Four Audio. The room adaption is quite different from PA system alignment, where equalizers are used to counteract room resonances. In the present application, room resonances are a major aspect of the natural room acoustic properties and therefore basically no PA system equalization for room correction should be applied.

The second portable source has omni-directional characteristics; it was designed and developed in the context of this thesis and an affiliated diploma thesis (Meuleman, 2011). It is called Sonic Ball and presented in (Meuleman et al., 2011). The Sonic Ball

system is depicted in Figure 4.17 (d). The mid/high unit of the source is a concentric dodecahedron with two shells. The inner shell works from 100 Hz up to approximately 3 kHz. It is equipped with 12 x 6.5" drivers and constructed like a standard dodecahedron loudspeaker commonly used for room acoustic measurements. Frequencies above 3 kHz are radiated by the outer shell that is equipped with 144 x 1" calottes, mounted in an aluminum framework and covered by a steel grille for protection. A dedicated subset of the calottes works with reversed polarity in order to suppress side lobes. Below 100 Hz, the system is supplemented by a 15" subwoofer that is located below the mid/high unit. The system is controlled and aligned using a DSP and driven by class-D power amplifiers. Due to the use of performant low- and mid-range drivers, combined with a large number of calottes, the system is capable of producing a sound power level of $L_w > 120$ dB with distortion factors $k < 1\%$ on its entire frequency range, which makes it suitable for measuring even large spaces with convenient SNRs. The radiation properties of the Sonic Ball fulfill the requirements defined in ISO 3382-1 (DIN, 2009). Furthermore, the source is capable of holding these properties up to much higher frequencies than specified in ISO 3382-1. The source was constructed for delivering high sound pressure levels, good omni-directional radiation properties, and a very flat energy frequency response. The latter is particularly important for auralization purposes. For more detailed information on the Sonic Ball speaker system, the reader is referred to (Meuleman et al., 2011). Sonic Ball was also used for the measurement in the SEACEN project (Weinzierl et al., 2012).

4.3.3.3 Receivers

The impulse responses were captured using several different receivers, cf. Figure 4.18. All receivers were accurately placed at a common pivot point. The VariSphear microphone array (cf. Section 4.1) was used to acquire array room impulse responses (ARIRs). Different grids and array configurations were captured using either an omni-directional Earthworks M30 microphone in a rigid sphere or a Microtech Gefell M900 microphone in an open sphere configuration. Additional omni-directional center impulse responses were captured at the pivot point, thus at the center of the array. Binaural room impulse responses (BRIRs) with 1° resolution in the horizontal plane were captured using a Neumann KU 100 dummy head (Norbert) mounted on the VariSphear motion system. Furthermore, some static classical stereophonic microphone configurations, such as AB, XY, ORTF and M/S, with Shoeps MK2, MK4, Neumann KM83, U89 and SM69 microphones were arranged for capturing basic stereophonic room impulse responses.

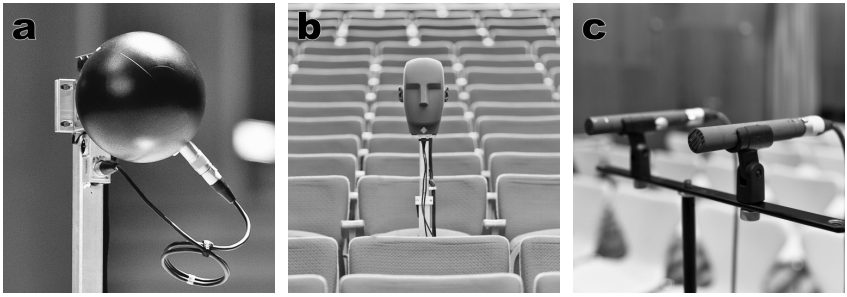


Figure 4.18 Receivers: a.) VariSphear scanning microphone array in rigid sphere configuration, b.) Motorized Neumann KU 100 dummy head (Norbert), c.) Exemplary stereophonic small-AB setup [Pictures: P. Stade].

4.3.3.4 Software, Hardware, and Audio Parameters

The impulse response measurements and motion control were both performed using the proprietary VariSphear array software running under MATLAB®. The impulse responses were acquired using sine-sweep excitation with emphasized (+20 dB low-shelf at 100 Hz) sine-sweeps of 2^{19} samples as proposed by Müller (1999). The audio interface was an RME Fireface UCX including its internal microphone preamplifiers.

4.3.3.5 Spherical panorama pictures and CAD models

In addition to the audio data, meta-data content, such as spherical panorama pictures or basic CAD models of the rooms, was generated. For capturing spherical panorama pictures, the VariSphear motion base was equipped with a custom built pano-head holding a Canon EOS 5D camera. Proprietary software for capturing and stitching the panorama pictures was developed by a students' work group from the media technology degree course at Cologne University of Applied Sciences (Stade et al., 2012). The capturing software is written in MATLAB® and the image stitching algorithm and viewer are written in C++. The packages is called GIXEL[4]. GIXEL enables the stitching and viewing of panorama images and the overlay of sound field data matrices from the SOFiA toolbox, cf. Section 4.2. This procedure can be used to visualize acoustic reflections in a panorama image of the surrounding room (Bernschütz et al., 2012).
The VariSphear array system was equipped with a laser head for capturing room dimensions and constructing basic true-to-scale CAD models referred to the measurement

[4] GIXEL: Giga-Pixel

Figure 4.19 VariSphear pano head with remotable Canon EOD 5D camera for capturing spherical panorama pictures [Picture: P. Stade].

origin. The VariSphear laser head and an exemplary CAD model are depicted in Figure 4.20.

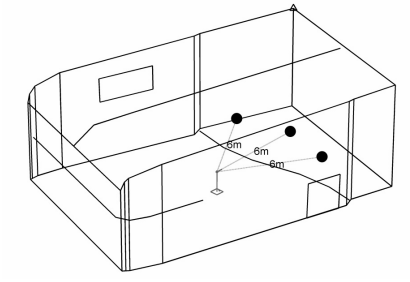


Figure 4.20 VariSphear laser head with remotable TOPCON EM-30 laser distance measuring device for capturing CAD models [Picture: S. Moritz] and exemplary basic CAD drawing of the small broadcast studio.

4.3.4 Headphone Equalization

In the context of acquiring the HRTFs from Section 4.3.2 and the BRIRs from Section 4.3.3, respective headphone compensation filters as discussed e.g. in (Schärer and Lindau, 2012) were generated for several headphone models. The headphones have each been positioned 12 times on the Neumann KU 100 dummy head for obtaining repre-

sentative transfer functions by averaging. Complex-phase FIR filters were generated using a semi-automatic inversion of the spline-based smoothed magnitude and phase responses of the mean of both median transfer functions for the left and the right ear in a next step. The lowest and highest frequency sections were excluded from the inversion, considering the physical limits of the drivers. Furthermore, high-Q dips in the responses were excluded from the inversion in order to avoid ringing artifacts of the filters.

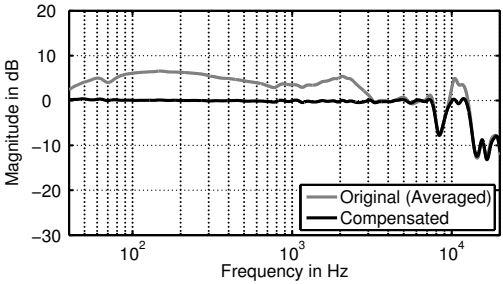


Figure 4.21 Headphone magnitude response compensation of the Neumann KU 100 dummy head and the AKG K 601 headphone. The gray curve shows the average for both ears and 12 repositions without compensation filters. The black curve shows the equalized curve with applied compensation filter.

The procedure for generating the FIR filters is similar to the approach for equalizing loudspeakers proposed in (Müller, 1999). The filters for several headphone models are published in the corresponding data sets from (Bernschütz, 2013a) and (Stade et al., 2012). For the listening experiments presented in Section 5, AKG K 601 headphones were used and the specific headphone filter was applied. In order to keep the latency low, which is of great importance in the context of dynamic binaural synthesis, the complex-phase FIR compensation filters were converted to minimum phase filters (2048 filter taps) using the Hilbert transform (Ohm and Lüke, 2004, pp 53–55). As a consequence, the compensation is restricted to equalizing the magnitude response of the dummy head-headphone chain. The original averaged magnitude of the Neumann KU 100 → AKG K 601 chain, and the compensated magnitude response are depicted in Figure 4.21. Nevertheless, for the presented experiments the use of compensation filters is not vital, since references were provided in all trials and only relative ratings were gathered. The chain of Neumann KU 100 and AKG K 601 delivers a quite satisfying transfer function on its own, due to the well-done internal diffuse field tuning of both the dummy head and the headphones, as proposed by Theile (1986b).

4.4 SCALE Software Tool

In order to perform listening experiments for this thesis and the associated research project, a proprietary listening test software, called Scale[5], was implemented and published by Vázquez-Giner (2013). In addition to the MUSHRA and SAQI designs that are presented in Section 5, the software includes several other test paradigms. Scale offers a subject administration database and a module for previewing listening test results. It communicates with the IOSONO Core renderer and the SoundScape Renderer (SSR) (Geier et al., 2008), (Geier and Spors, 2012) for performing spatial audio listening experiments (Vázquez-Giner, 2015). Screen shots of the Scale user interface are depicted in Figure 5.2 and Figure 5.3. A comparable tool from Technical University of Berlin, called WhisPER, is described in (Ciba et al., 2009).

4.5 Environment for the Listening Experiments

All listening tests presented in this thesis were performed in the anechoic chamber at Cologne University of Applied Sciences. This environment offered low background noise levels and minimized distraction of the participants. The subjects were sitting in a swivel chair that enables full 360° rotation, wearing tracked headphones and holding a tablet computer for operating the Scale software interface. The environment is depicted in Figure 4.22.

Figure 4.22 Listening experiment environment – a participant is sitting in the swivel chair in the anechoic chamber performing a MUSHRA test [Pictures: P. Stade].

[5] Scale – **S**etup, **c**onduction, and **a**nalysis of **l**istening **e**xperiments

4.6 Technical Setup for the Listening Experiments

The technical setup for performing the listening experiments is depicted in Figure 4.23. The experiments are controlled by the Scale software (Section 4.4) that is running on a standard office PC. Scale controls the IOSONO Core renderer via TCP/IP and provides the anechoic and diotic audio feeds via AES3. The IOSONO Core renderer holds the spatial filters and performs the fast real-time convolution (buffersize 256 taps) of the audio feed and the spatial filters, taking into account the subject's head rotation. The subject's head rotation is captured using a Polhemus Fastrak tracking device that is connected via RS232 $\leftrightarrow$ RS485 $\leftrightarrow$ RS232 to the IOSONO Core. Only horizontal rotations (yaw) are considered for the experiments. The resolution of the spatial filters is 1° in the horizontal plane for all scenarios and stimuli. The audio signal is converted and amplified using a RME ADI-2 digital to analog converter and played back using AKG K601 headphones. A wireless tablet computer (Apple iPad 2) with touch surface mirrors the Scale user interface via VNC and serves as user input device to perform the ratings, cf. Figure 4.22.

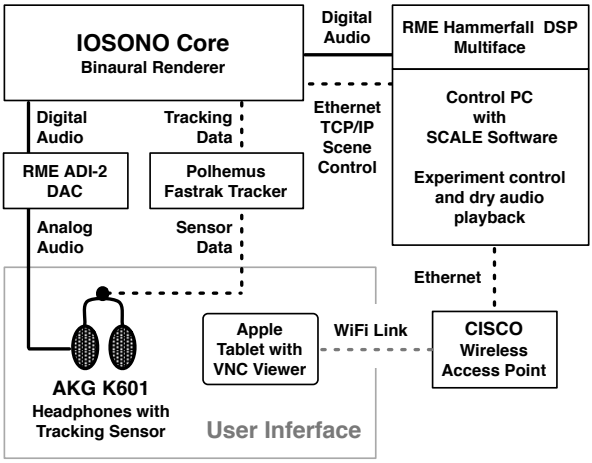


Figure 4.23 Technical setup for the listening experiments.

5 Listening Experiments

Listening experiments were conducted in order to assess perceptual influences of certain physical or algorithmic parameters and to evaluate the methods that are presented in this thesis.

5.1 Two-stage Approach and Test Paradigms

The experiments were performed in a two-stage approach using two different test designs. A multiple stimulus with hidden reference and anchor (MUSHRA) paradigm, refer to Section 5.1.1 and (ITU, 2003), was used to assess the quantitative influence of a physical or algorithmic parameter on the unspecific overall perceived quality in the first stage. In the second stage, selected representative stimulus-reference pairs were picked out for assessing specific perceptual attributes based on the spatial audio quality inventory (SAQI), refer to Section 5.1.2 and (Lindau et al., 2014). The two-stage approach is illustrated in Figure 5.1.

5.1.1 MUSHRA

The multiple stimulus with hidden reference and anchor (MUSHRA) test is a double-blind multi-stimulus test method with hidden reference and hidden anchor(s) proposed by the International Telecommunication Union (ITU) in Rec. ITU-R BS.1534-1 for the «subjective assessment of intermediate quality of coding systems» (ITU, 2003). Whereas the word «subjective» should not be misunderstood in this context. The MUSHRA test claims to deliver objective results, as can be expected from any reasonable (listening) test design (Weinzierl and Maempel, 2012).

The MUSHRA paradigm is primarily designed for assessing the perceptual audio quality of lossy audio transmission systems and codecs. During the test procedure, several different stimuli or stimulus levels are presented at once. A stimulus can be compared to a given reference and to other stimuli. Amongst the presented stimuli, at least one hidden reference (HR) and one anchor (AN) are presented. The ratings are gathered on a continuous ordinal scale {*bad, poor, fair, good, excellent*} combined with a ratio scale (0-100%), which is referred to as continuous quality scale.

The test is recommended for intermediate quality systems, i.e. some of the presented stimuli besides the anchor are assumed to show artifacts or changes that are clearly

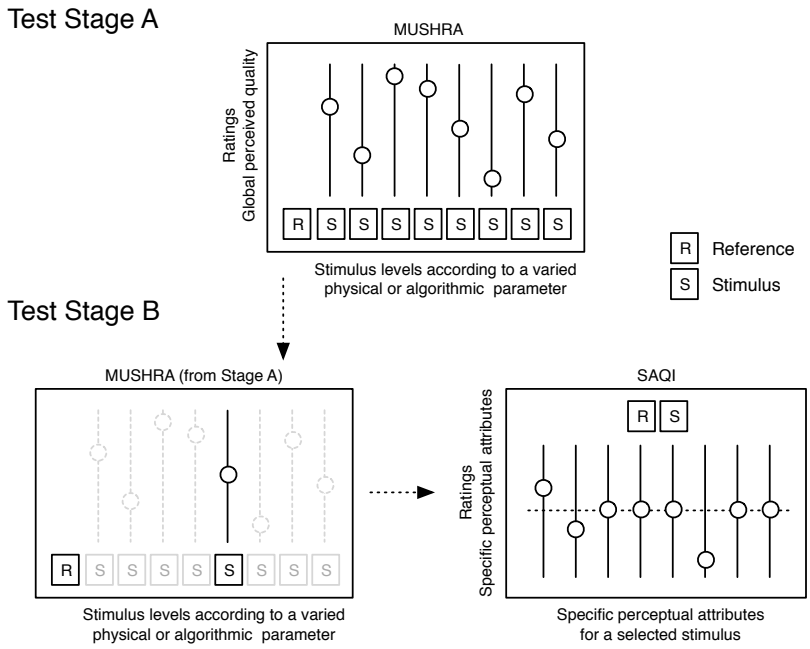


Figure 5.1 Illustration of the two-stage approach for the listening tests.

perceivable by the test subjects. If the occurring artifacts for all or most of the stimuli are more subtle, a different test design such as an ABX test as proposed in Rec. ITU-R BS.1116-1 (ITU, 1994) is preferable. The MUSHRA test design is claimed to deliver statistically significant results even for a comparatively low number of ratings (ITU, 2003). Nevertheless, the test might be prone to a certain bias (Zielinski et al., 2007) (Zielinski et al., 2008). All in all, the procedure allows for evaluating comparatively large stimulus sets with low time exposure. This is reasonable in the given context, as the tests aim to give a first and broad exploratory overview on the absolute magnitude of influence of widely varied physical or algorithmic parameters on the perceived audio quality.

The MUSHRA tests in this thesis are largely based on the original recommendation (ITU, 2003), except for using dynamic binaural synthesis instead of static stimuli, and, therewith, using different anchors than proposed in the document. The source signal, cf. Section 5.5, was constantly looped and the subject switched over different spatial

filters in the binaural rendering engine in real-time, which allowed for immediate direct comparison of the stimuli. The subjects were asked to rate the audio quality in terms of the magnitude of the overall perceived difference compared to the given reference. This is assumed to provide the best imaginable quality, even though neither particularly high plausibility nor authenticity is claimed for any of the reference stimuli. The dynamic behavior of the stimuli, i.e. their dependence on the wave incidence direction and, thus, the relative rotation of the subject referred to the virtual source turned out to be a particular challenge. The subjects were encouraged to perform head and body rotations during the rating procedure and to give averaged judgments. This additional degree of freedom turned the procedure into a demanding task for the participants.

In the meantime, an updated proposal for the MUSHRA design is available (ITU, 2014), which includes more specific details on the test and evaluation procedure. However, the tests presented in this thesis are still based on the previous proposal (ITU, 2003).

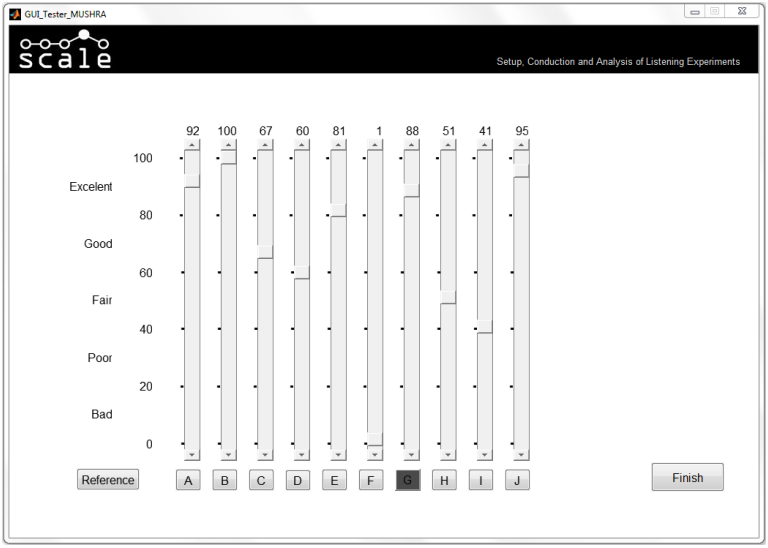


Figure 5.2 Exemplary MUSHRA test interface in *Scale* (Vázquez-Giner, 2013) for 8 different stimuli plus hidden reference and anchor.

5.1.2 SAQI

The spatial audio quality inventory (SAQI) is a consensus vocabulary consisting of verbal descriptors of perceptual attributes for assessing specific apparatus-related perceptual differences between different virtual auditory environments (VAEs), and between VAEs and either a presented or imagined reality. SAQI was developed and proposed by Lindau et al. (2014). A proposal for instrumentalization is given by the same research group in a 2014 release (Ciba et al., 2014) of the WhisPER toolbox (Ciba et al., 2009). The SAQI test is a semantic differential test paradigm (Bortz and Döring, 2002, pp 185–187) using the SAQI descriptors combined with bipolar, unipolar, or dichotomous rating scales, which individually depends on the specific attribute.

The SAQI test procedure is implemented in the Scale software, refer to Section 4.4. For a paired comparison, the subject can seamlessly toggle between two different spatial filters while rating. The SAQI test interface in Scale is depicted in Figure 5.3. Since only a single source was presented and not the entire technical VAE system but only the presented stimulus pairs themselves were to be evaluated, the optional SAQI assessment entities and the aspects of time variance were ignored. The SAQI test is well-suited for a more differentiated insight into the information gathered using the MUSHRA tests. While the MUSHRA test is used for obtaining quantitative and unspecific perceptual quality ratings for a broad range of different stimuli, the SAQI test is used for extracting detailed information on specific perceptual attributes for selected stimulus-reference pairs in a second stage.

The subset of attributes from the SAQI vocabulary used for the listening experiments is listed in Table 5.1. The SAQI tests were performed using the German version (SAQI-GER) of the vocabulary, as the SAQI participants were native German speakers. Table 5.1 and the SAQI plots show the corresponding validated English translations (SAQI-EN). For a back translation to the original German attribute names, and for detailed circumscriptions refer to the SAQI test manual (Lindau, 2014b).

Table 5.1 Listing of the involved SAQI attributes.

Catergory	Perceptual Attribute	Scale end labels
Difference	Difference	none-very large
General	Clarity	less pronounced-more pronounced
	Naturalness	lower-higher
	Presence	lower-higher
	Degree of liking	lower-higher
Timbre	Tone color dark-bright	darker-brighter
	High-frequency tone color	attenuated-emphasized
	Mid-frequency tone color	attenuated-emphasized

Table 5.1 Listing of the involved SAQI attributes.

Catergory	Perceptual Attribute	Scale end labels
	Low-frequency tone color	attenuated-emphasized
	Sharpness	less sharp-sharper
	Roughness	less rough-more rough
	Comb filter coloration	less pronounced-more pronounced
	Metallic tone color	less pronounced-more pronounced
Tonalness	Tonalness	more unpitched-more pitched
	Pitch	lower-higher
Dynamics	Loudness	quieter-louder
	Dynamic range	smaller-larger
	Dyn. compression effects	less pronounced-more pronounced
Geometry	Distance	closer-more distant
	Depth	less deep-deeper
	Width	less wide-wider
	Height	less high-higher
	Externalization	more internalized-less internalized
	Localizability	more difficult-easier
	Spatial disintegration	more coherent-more disjointed
	|Horizontal direction|[1]	not shifted-shiftet (max. 30°)
	Vertical direction	shifted up-shifted down (min.−90°, max. +90°)
Room	Reverberation level	less-more
	Reverberation time	shorter-longer
	Envelopment	less pronounced-more pronounced
Time	Pre-echoes	less intense-more intense
	Post-echoes	less intense-more intense
	Temporal disintegration	more coherent-more disjointed
	Crispness	less pronounced-more pronounced
Artifacts	Pitched artifact	less intense-more intense
	Impulsive artifact	less intense-more intense
	Noise-like artifact	less intense-more intense
	Ghost source	less intense-more intense
	Distortion	less intense-more intense
Other	Other[2]	less pronounced-more pronounced

Once again, the dynamic behavior of the stimuli, depending on the rotation of the subjects referred to the virtual source, made the procedure very demanding for the

[1] In contrast to the original proposal, an absolute value of the horizontal shift is assessed instead of clock- and counterclockwise shifts, since some of the stimuli showed ambiguous shifts depending on the respective subject's head or body rotation referred to the virtual source.

[2] "Other" is a mandatory residual category/attribute introduced to catch any perceptual artifact that might not be covered within the current SAQI vocabulary. During the experiments that were conducted in this thesis, none of the participants made use of this residual category.

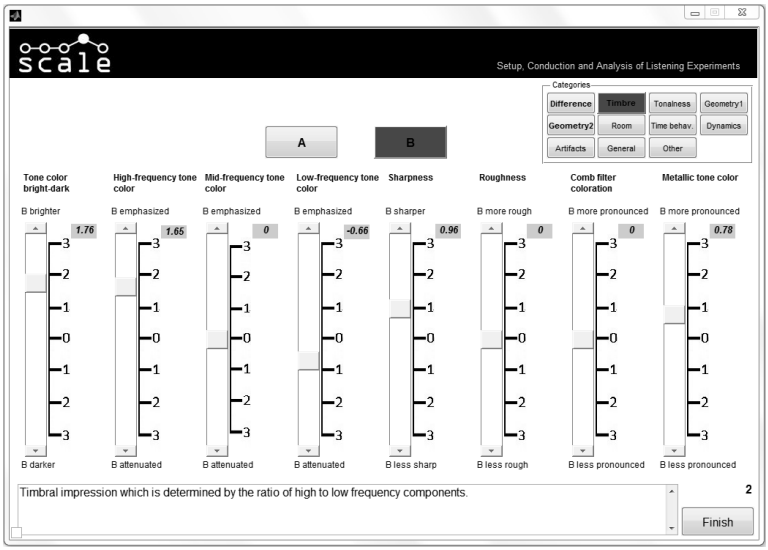


Figure 5.3 Exemplary SAQI test interface in Scale (Vázquez-Giner, 2013) for a paired comparison. In the upper center section the stimulus A or B can be switched seamlessly in real-time. In the upper right corner the SAQI categories {difference, timbre, tonalness, geometry, room, time behavior, dynamics, artifacts, general impressions, other} can be selected and the faders adapt accordingly. In the bottom box the circumscription for a specific attribute is shown when touching the respective fader.

participants. The rotational orientation entailed a dynamic variation of several perceptual attributes. The participants were instructed to perform head and body rotations and to give best possible average ratings.

5.2 Statistical Evaluation

The majority of tests conducted in this thesis aim to assess and directly reflect the subjects' comparative judgments for certain stimulus pairs or stimulus clusters instead of being strict statistical hypothesis tests. Thus, most of the information for both test paradigms can directly be deduced from the corresponding plots. The plots show the rating means with their respective confidence intervals (CIs). For the calculation of the confidence intervals a bootstrapping approach (DiCiccio and Efron, 1996) using $2 \cdot 10^3$ bootstrap samples at a confidence level of 95% is applied. Bootstrap CIs are based on

resampling the underlying datasets and are claimed to be more accurate than classical confidence intervals (DiCiccio and Efron, 1996). The upper and lower interval limits are evaluated separately and therefore the CIs might turn out to be asymmetric.

A one-way analysis of variance (ANOVA) (Thompson, 2006, pp 303–224) paired with a Tukey-Kramer post hoc test (Thompson, 2006, pp 325–332) at a confidence level of 95% is applied in order to evalute the statistical significance of difference between dedicated stimuli. Thus, whenever the rating means are supposed to be equal or the null hypothesis is rejected and statistical significance is claimed for the means of a stimulus pair or certain stimulus clusters, the statements are based on the results of this procedure. Sporer et al. (2009) propose alternative and extended approaches for the statistical evaluation of MUSHRA tests, where they consider that the subjects performing inherent ranking tests and a paired comparison test between pairs of stimuli. In this thesis the statistical evaluation is restricted to the classical procedure stated above. The calculation of means and boostrap confidence intervals, as well as the ANOVA and post hoc tests are performed using the MATLAB® statistics toolbox.

5.3 Participants/Grading

For different listening tests designs the required experience and expertise of the participating subjects is considerably different. While some of the basic MUSHRA tests can be performed by adequately trained but otherwise relatively inexperienced subjects, the SAQI tests generally demand for well-trained and highly experienced participants with high audio expertise.

Three groups of participants from different resources were involved. They were graduated into different categories, cf. Table 5.2. For less critical tests, a large mixed group of university students in the field of media technology without specific listening experience or specific audio expertise was recruited. This group is referred to as normal listeners and assigned to category C. A second more specialized group with certain minimum requirements was assembled for performing more critical tests. The participants of this group had to have enhanced listening experience (e.g. audio mixing or playing musical instruments) on an ambitious but not necessarily professional level. They had to commit to regular participation in all respective listening experiments. This group is referred to as experienced listeners and assigned to category B. The third category is a carefully selected group of audio experts with extensive experience and professional expertise. All members of this specific group were required to have several years of experience and to work full-time in the field of audio engineering. The group is composed of sound engineers from the Westdeutscher Rundfunk (WDR) radio broadcast studios, some mixed external experts and selected members of the audio research group

at Cologne University of Applied Sciences. This group is referred to as expert listeners and assigned to category A. An informal post-hoc analysis indeed showed particularly plausible and consistent ratings from all listeners in category A.

Table 5.2 Grading of the participants and respective requirements/expertise.

Grading	Requirements/Expertise
Category C (Normal listeners)	Good mood No hearing damage Taking part in the introduction and training
Category B (Experienced listeners)	All from category C plus: Audio and/or listening experience on an ambitious but not necessarily professional level Constantly taking part in all listening experiments
Category A (Expert listeners)	Professional audio expertise Several years of experience

5.4 Introduction and Training

For the MUSHRA tests, all subjects obtained around 40 minutes of general introduction to the experiments. Additionally, every subject was provided a personal explanation by an adviser before starting the experiment. Before starting a scenario, a detailed pre-produced verbal explanation of the ensuing question, including a description of the scale end points and audio examples, was presented to each subject. All subjects had at least one hour of listening experience using the binaural rendering system and handling the test environment in hidden training sessions. Those are the minimum requirements for category-C listeners.

For the SAQI tests, each participant obtained an individual introduction for around 1.5 h; the introduction included a dedicated explanation and discussion of the single SAQI attributes.

A short audio drama including spatially distributed conversation, effect sounds, and a short piece of Flamenco guitar, virtually reproduced in a broadcast studio of the Westdeutscher Rundfunk (WDR), was produced and played back in advance to the very first test procedure for motivating the subjects. Without exception the participants

gave positive feedback for the teaser and indeed reported a considerable increase in personal motivation.

Only very few single trials or subjects needed to be excluded from the statistical evaluation due to obviously inappropriate ratings. This indicates that the participants were taking their task seriously. It also showed that the overall test procedure, the stimulus ranges, and the introduction and training were well designed and adequate.

5.5 Source Signal

The source signal is an anechoic monaural audio signal. This signal is routed through the binaural renderer that applies the spatial filters and produces the binaural signal. The combination of source signal and spatial filter is referred to as stimulus in this thesis. Different types of source signals were evaluated in informal pre-tests performed by three category-A listeners. Pure technical audio stimuli, such as noise bursts or clicks, as well as natural stimuli, such as drums, guitars, speech, singing vocals, or violins, were evaluated. For the MUSHRA test, the use of critical material representing the typical broadcast program for the desired application is recommended in (ITU, 2003). This recommendation can only partly be transferred to the given test scenarios, since typical material can hardly be defined in this context. Pure technical stimuli were excluded due to their questionable relation to realistic signals.

A drum loop (kick, snare and hi-hats) turned out to be highly critical with respect to the artifacts and to be comfortable for listening during longer sessions. It offers a broad frequency range, which is useful for revealing spectral artifacts, and sharp transients for the detection of temporal artifacts. This stimulus is referred to as Drums and it is consistently used throughout all listening experiments except MUSHRA-STM and MUSHRA-BCS, where the influence of the source signal itself is evaluated or a less critical scenario is presented. The alternate source signal for the latter is an acoustic guitar with steel strings and rich high-frequency content, which can be considered as critical but not highly critical. The alternate source signal is referred to as Guitar. Other stimuli, such as speech or singing vocals, as well as some classical bowed instruments or classic guitar, were found to be considerably less critical in this specific context.

5.6 Simulated Data Sets

The listening experiments are divided into experiments with simulated or measured data sets. Simulated data sets are analytic array responses based on mathematical wave descriptions. Measured data sets are microphone array recordings or measurements using a sound source in a room. When analyzing measured data sets, the influences of particular parameters cannot be evaluated separately, since several factors of influence inextricably emerge at the same time. Therefore, simulated data sets are used in a first step, which allow for an independent selective analysis of a particular parameter while maintaining ideal conditions for all other parameters. Listening experiments based on simulated data sets are presented in this section. The system architecture for generating the respective data sets is described in Figure 5.4.

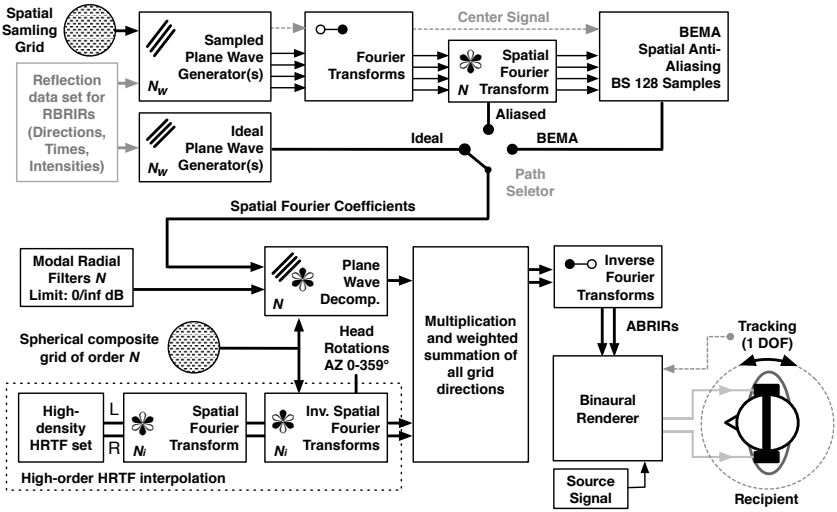


Figure 5.4 System architecture for generating the simulated data sets and presenting the stimuli.

5.6.1 Modal Reduction

One of the major issues concerning array based binaural auralization is the limited modal resolution of microphone arrays, which is discussed in Section 3.5. Different aspects concerning sound field descriptions with reduced modal order and their impact on perception are analyzed in the following. Full resolution HRIR or BRIR sets are

compared to their respective counterparts with limited modal resolution. They are denoted as head-related impulse response with limited modal resolution (RHRIR)[3] and binaural room impulse response with limited modal resolution (RBRIR). Both are generated using simulated plane waves with reduced modal order that are adapted to the HRIR set as described in Section 3.5 and depicted in Figure 5.4. The underlying theoretical aspects, as well as the listening experiments presented in this section are published in Bernschütz (2014) and in Bernschütz et al. (2014).

5.6.1.1 MUSHRA-MR I

In Section 3.5, it is shown that the properties of the RHRTF depend on the approach of adaptation between the spherical HRTF set and the (order-limited) modal wave field description. It is shown that an inconvenient adaptation entails severe low-pass filter effects due to the modal mismatch of HRTFs and low-order sound field descriptions. The properties of the RHRTFs can be considerably improved by performing spatial downsampling of the HRTFs with a composite grid that exactly corresponds to the wave order. Furthermore, the properties not only depend on the order but also on the node distribution of the composite grid. In the first experiment MUSHRA-MR I, three adaptation approaches are directly compared in a single trial for different low wave orders. The first variant involves a high-order composite grid (Gauss, $N = 35$) that enables nearly ideal resolution of the HRTF set in the audible frequency range. Using a high-order composite grid corresponds to spatial upsampling of the low-order wave description. This adaptation entails maximum low-pass filter effects due to the order mismatch. Both remaining variants involve low-order composite grids that are adapted to the wave order for two different quadrature types, Gauss and Lebedev quadratures, see Section 3.2.2. The experiment parameters are listed in Table 5.3 and the result is depicted in Figure 5.5.

The results of MUSHRA-MR I reflect the effects of the predicted low-pass effect for the high-order grid; all mean ratings are located within the lowest third of the rating scale. Additionally, the expected shift of the low-pass filter knee towards higher frequencies with increasing order can be observed. Higher orders entail higher ratings. At sufficiently high orders (i.e. N>35), the low-pass vanishes from the audible range. Both versions of composite grids that are adapted to the lower wave order show clear and statistically significant improvements for the presented orders. The order-adapted Gauss and order-adapted Lebedev quadratures achieve different ratings. A clear perceptual enhancement using the Gauss grid can be approved for the presented orders. This outcome is surprising, since the mathematical mean overall spectral error

[3] The respective time-frequency Fourier transform is denoted as RHRTF.

Table 5.3 Parameters MUSHRA-MR I.

Experiment ID	MUSHRA-MR I (Modal Reduction)
Test paradigm	MUSHRA
Varied parameter(s)	Plane wave with limited modal orders $N = 3 - 6$ for three different wave adaptation types: High-order, adapted Gauss and adapted Lebedev
Spatial filter(s)	RHRIRs
Reference	HRIRs
Anchor	Unfiltered diotic
Number of ratings	22 (hidden double trials)
Listener grading	Mixed B(6), A(5)
Trials	1
No. of stimuli per trial	12 + Anchor + Hidden Reference
Source signal	Drums

between RHRTFs and HRTFs is a little lower for the Lebedev grid than for the Gauss grid, c.f. Section 3.5. However, at least for the presented scenario with a single frontal source and the first few orders, the Gauss grid shows a better performance throughout from a perceptual point of view. Finally, the RHRTFs converge to the original full resolution HRTFs with increasing orders for all approaches. At low modal orders, clearly significant physical and perceptual improvements can be observed for order-adapted composite grids.

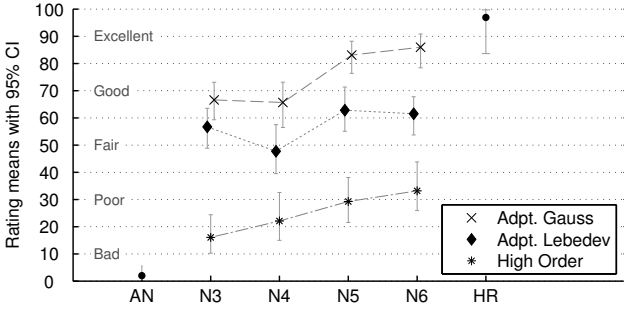


Figure 5.5 MUSHRA-MR I results.

5.6.1.2 MUSHRA-MR II

The experiment MUSHRA-MR II is similar to MUSHRA-MR I. The stimuli using the high-order composite grid are dropped and the range of order-adapted stimuli is extended up to order $N = 8$ within one trial. The experiment was performed with the same listeners as MUSHRA-MR I. The experiment parameters are listed in Table 5.4 and the results are depicted in Figure 5.6.

Table 5.4 Parameters MUSHRA-MR II.

Experiment ID	MUSHRA-MR II (Modal Reduction)
Test paradigm	MUSHRA
Varied parameter(s)	Plane wave with limited modal orders $N = 3 - 8$ for two different composite grids
Spatial filter(s)	RHRIRs
Reference	HRIRs
Anchor	Unfiltered diotic
Number of ratings	22 (hidden double trials)
Listener grading	Mixed B(6), A(5)
Trials	1
No. of stimuli per trial	12 + Anchor + Hidden Reference
Source signal	Drums

For orders $N = \{3, 4, 5, 6\}$, MUSHRA-MR II shows similar results as MUSHRA-MR I just with a typical context-dependent bias. The additional orders $N = 7$ and $N = 8$ show two interesting aspects. First, a positive "outlier" for $N = 7$ using the Lebedev quadrature emerges. This indicates a volatile behavior of the Lebedev quadrature and a more constant performance of the Gauss quadrature over different orders. These properties further consolidate in the following experiments. For some particular orders (e.g. for $N = 3$ and $N = 7$), the Lebedev quadrature apparently delivers similar perceptual results as the Gauss quadrature. Second, the Gauss grid for $N = 8$ delivers a mean rating that surpasses the 90%-excellent boundary for a critical stimulus and critical listeners of the categories A and B only. A clearly perceivable but relatively small difference to the reference remains. MUSHRA-MR II covers a realistic range of orders that can be resolved by typical microphone arrays according to the current state of technology.

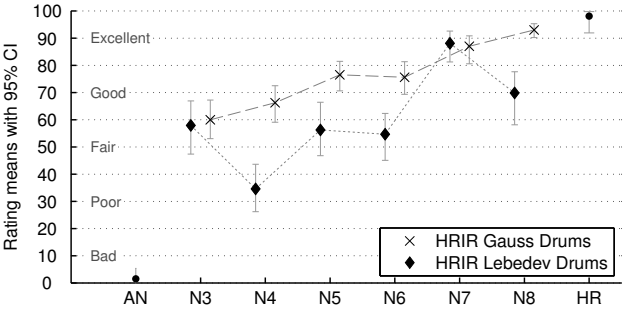


Figure 5.6 MUSHRA-MR II results.

5.6.1.3 MUSHRA-MR III

MUSHRA-MR III is again similar to the previous experiments. However, a broader range of orders is presented, i.e. $N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 17, 23\}$. Due to the large number of stimuli, the experiment is split into two different trials (Gauss, Lebedev). In contrast to the previous experiments, MUSHRA-MR III involves several additional category-C listeners. The hidden double trials were dropped due to the larger number of participants. The experiment parameters are listed in Table 5.5 and the results are depicted in Figure 5.7.

Table 5.5 Parameters MUSHRA-MR III.

Experiment ID	MUSHRA-MR III (Modal Reduction)
Test paradigm	MUSHRA
Varied parameter(s)	Plane wave with limited modal orders $N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 17, 23\}$ for two different composite grids
Spatial filter(s)	RHRIRs
Reference	HRIRs
Anchor	Unfiltered diotic
Number of ratings	26 (each trial)
Listener grading	Mixed C(15), B(6), A(5)
Trials	2
No. of stimuli per trial	13 + Anchor + Hidden Reference
Source signal	Drums

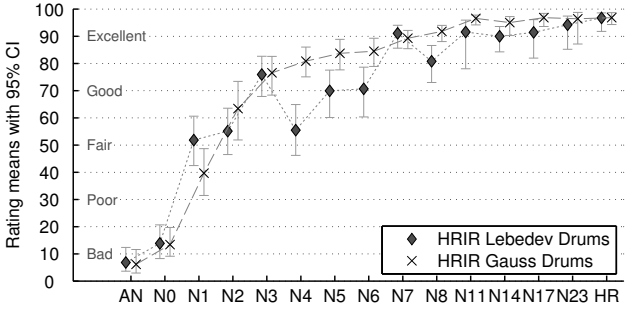


Figure 5.7 MUSHRA-MR III results.

The ratings for the previously presented orders are similar to the ratings from MUSHRA-MR I and MUSHRA-MR II, even though a considerable number of less critical category-C listeners is involved. Using the Gauss grid, ratings from order $N = 11$ upwards do not show statistically significant differences to the reference anymore. The volatile characteristics of the Lebedev grid and the smooth and more constant performance of the Gauss grid are shown again. All MUSHRA-MR experiments show negative outliers for the Lebedev $N = 4$ stimulus. The processing chain and stimuli were double checked for potential errors. The Lebedev quadrature apparently tends to perform better on odd orders than on even ones.

5.6.1.4 SAQI-MR

For SAQUI-MR the selected stimulus is a RHRTF set based on a Gauss composite grid with order $N = 5$. This specific stimulus yielded ratings in the range of 75-85% (between good and excellent) in the MUSHRA-MR experiments. Thus, the stimulus is still acceptable but shows clearly audible differences compared to the reference, which is a native HRTF set with full modal resolution. The selected SAQI attributes are listed in Table 5.1 and the experiment parameters in Table 5.6.

The results are depicted in Figure 5.8. The difference between stimulus and reference is clearly perceivable. For all general attributes, i.e. clarity, naturalness, presence, and degree of liking, the stimulus is rated lower than the reference. The stimulus appears to be perceived with a certain increase of mid/high frequencies, accompanied by increased comb-filter artifacts and a more metallic tone color. The listeners perceive a small dynamic compression effect. The source depth and width appear to be slightly increased, whereas the height remains identical. A horizontal shift of the source and a small vertical shift are perceived. The perceived distance and externalization appear to be comparable, at least concerning the overall mean ratings. However, the subjects

Table 5.6 Parameters SAQI-MR.

Experiment ID	SAQI-MR (Modal Reduction)
Test paradigm	SAQI
Varied parameter(s)	Plane wave with limited modal order $N = 5$ using a gauss composite grid
Spatial filter(s)	RHRIRs
Reference	HRIRs
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	Drums

reported a sudden reduction of externalization and varying distances for a few specific spatial orientations. The localizability clearly decreases and spatial disintegration of the source is perceived. The reverberation level and envelopment are rated to be slightly increased. These results appear to be confusing at first, since both the stimulus and the reference were completely dry and comprised no reverberation at all. But increased spatial and temporal disintegration, decreased localizability, and the subtle perception of a ghost source appears to be interpreted as additional reverberation.

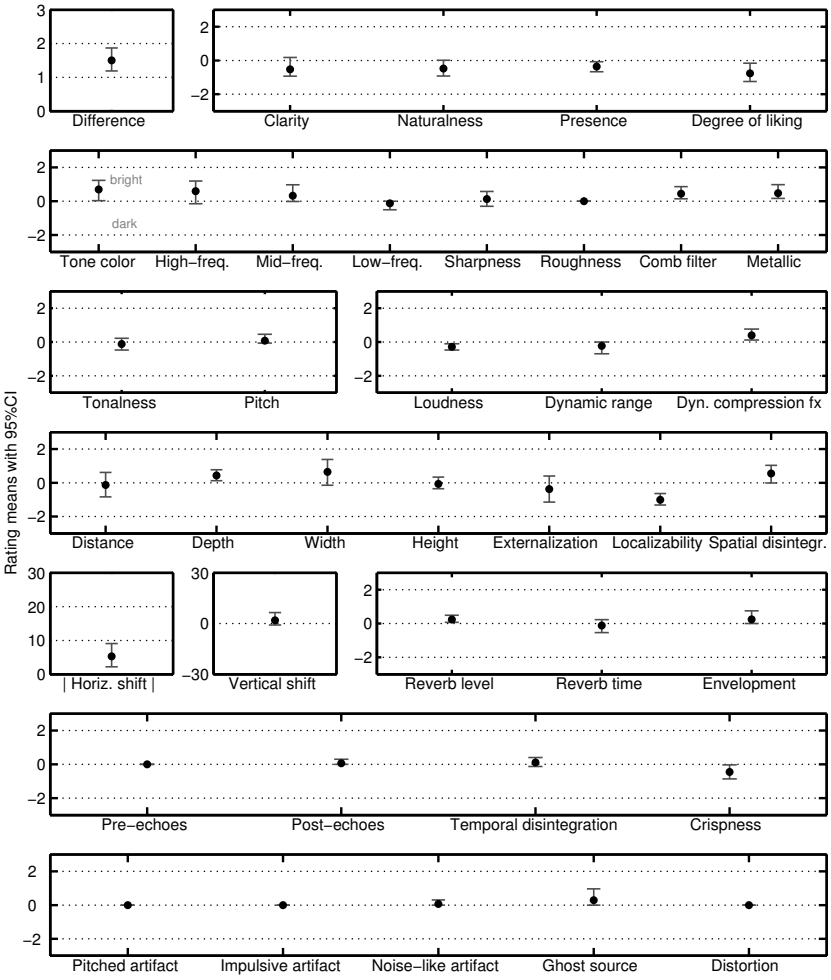


Figure 5.8 SAQI-MR results.

5.6.1.5 Early Reflections (MUSHRA-REF)

The next scenario is based on the MUSHRA-MR experiment. In order to analyze the influence of additional room reflections in addition to the direct sound, some early reflections according to the small broadcast studio at WDR Cologne (SBS) of the WDR are presented. A specific artificial BRIR set was generated for this purpose, which is described in the following.

5.6.1.5.1 Generating Simulated BRIRs In order to enable a well-controlled modal reduction that is free of different influences and effects, a specific BRIR set needs to be generated. The direct use of a measured array response is not suitable in this case, since a variety of unwanted side-effects arises. The simulated BRIR sets are implicitly based on microphone array measurements. An interposed procedure for appropriate signal abstraction is applied. The underlying microphone array data was captured at a broadcast studio (SBS), refer to Section 4.3.3. An algorithm provided in the SOFiA toolbox is used for extracting the directions of incidence Ω_k, the times of incidence τ_k, and the levels $\hat{p}_k$ in a high-mid frequency band around 2–4 kHz for $K = 24$ early reflections k within the first 200 ms of the room response. Simulated BRIRs $b^{\mathrm{l,r}}$ are generated from this basic geometric description by summing time-shifted and intensity-scaled high-order interpolated HRIRs $h^{\mathrm{l,r}}$ according to

$$b^{\mathrm{l,r}}(\Omega_{\mathrm{h}}) = \sum_{k=1}^{K} \hat{p}_k \, h^{\mathrm{l,r}}(\Omega_k - \Omega_{\mathrm{h}}, t - \tau_k), \tag{5.1}$$

where Ω_h describes the subject's head orientation. The identical geometric description is used for generating the respective order-reduced RBRIR sets using the plane wave generators as depicted in Figure 5.4. This approach enables perfectly isolated analysis of the perceptual influences that are introduced by a reduction of the modal order, even though the spectral properties of the corresponding measured BRIR set are substantially simplified. The respective simulated BRIR sets were evaluated and rated as realistic and well-sounding by three category-A listeners during informal listening sessions prior to the experiment.

5.6.1.5.2 MUSHRA-REF The experiment MUSHRA-REF extends the MUSHRA-REF III experiment and is designed for analyzing the perceptual influence of additional and likewise order-reduced reflections. The spatial filters are generated as discussed in the last section. Simulated BRIRs were presented for the Gauss and the Lebedev composite grid. The experiment parameters are listed in Table 5.7. The results, including the ratings from MUSHRA-REF III (dry HRIRs) for comparison, are depicted in Figure 5.9 and Figure 5.10.

Table 5.7 Parameters MUSHRA-REF.

Experiment ID	MUSHRA-REF (Modal red. with early reflections)
Test paradigm	MUSHRA
Varied parameter(s)	Plane waves (including reflections) with limited modal orders $N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 17, 23\}$ for two different composite grids
Spatial filter(s)	Simulated RBRIRs
Reference	Simulated BRIRs
Anchor	Unfiltered diotic (including reflections)
Number of ratings	26 (each trial)
Listener grading	Mixed C(15), B(6), A(5)
Trials	4 (2 trials from MUSHRA-REF III for comparison)
No. of stimuli per trial	13 + Anchor + Hidden Reference
Source signal	Drums

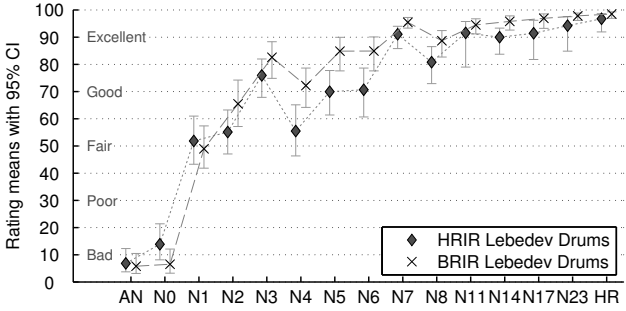


Figure 5.9 MUSHRA-REF results (Lebedev trials).

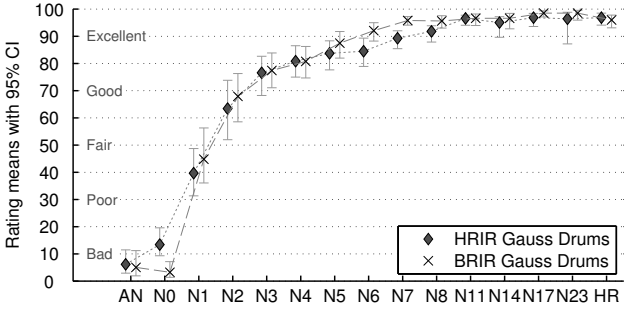


Figure 5.10 MUSHRA-REF results (Gauss trials).

The plots show an increase in the ratings for the (R)BRIRs compared to the dry (R)HRIRs with a Lebedev composite grid. For the Gauss grid, a slight increase can still be observed for some orders, even though the increase is more subtle. In general, additional reflections, even if reduced in the modal resolution, appear to mask the perceived order-reduction artifacts of the direct sound and tend to improve the perceptual properties of the signals, instead of introducing further audible artifacts. Diffuse reverberation, which is less prone to order-reduction artifacts due to the multitude of randomly distributed directions of incidence, is expected to increase the ratings due to other masking effects. The synthesis of highly diffuse and natural sounding reverberation with controllable modal resolution is a complex and time-consuming task. As a consequence, diffuse reverberation is considered for measured data sets only, refer to Section 5.7.

5.6.1.6 Source Signal (MUSHRA-STM)

The MUSHRA-STM experiment is a clone of the MUSHRA-MR III experiment, with a different source signal. The modified experiment is conducted in order to find out to what extent the perceptibility of the order-reduction artifacts depends on the source signal.

The highly critical drum source signal (Drums) is compared to a source signal with clearly different characteristics. The alternate stimulus is a western acoustic guitar (Guitar), which is rated as critical but not highly critical, refer to Section 5.5. The experiment parameters are listed in Table 5.7. The results, including the ratings from MUSHRA-REF III (source signal: Drums) for comparison, are depicted in Figure 5.11 and Figure 5.12.

Indeed, a less critical source signal entails slightly increased ratings, at least for certain orders of the Lebedev composite grid that comprises stronger perceptual reduction artifacts in general. For the Gauss grid, the differences are only subtle and not statistically significant. Thus, the experiment shows that the reduction artifacts are generally well perceivable and the corresponding ratings are comparable, even though an entirely different source signal is presented.

Table 5.8 Parameters MUSHRA-STM.

Experiment ID	MUSHRA-STM (Source Signal)
Test paradigm	MUSHRA
Varied parameter(s)	Plane wave with limited modal orders
	$N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 17, 23\}$
	for two different composite grids
Spatial filter(s)	RHRIRs
Reference	HRIRs
Anchor	Unfiltered diotic
Number of ratings	26 (each trial)
Listener grading	Mixed C(15), B(6), A(5)
Trials	4 (2 trials taken from MUSHRA-REF III for comparison)
No. of stimuli per trial	13 + Anchor + Hidden Reference
Source signal	Guitar

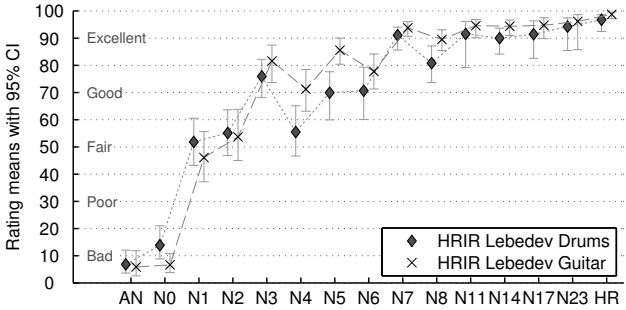


Figure 5.11 MUSHRA-STM results (Lebedev trials).

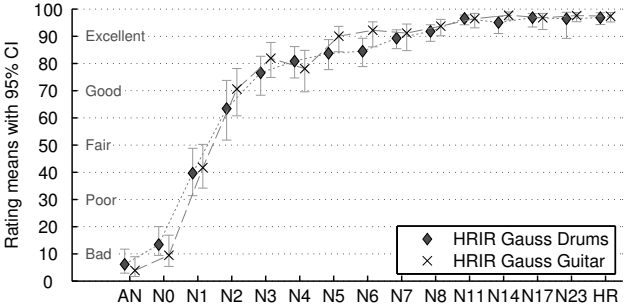


Figure 5.12 MUSHRA-STM results (Gauss trials).

5.6.1.7 Simulated Best-Case Scenario (MUSHRA-BCS)

The last experiment in the series of investigations of modal order-reduction based on simulated signals is MUSHRA-BCS, where a less critical scenario is presented that is supposed to be close to realistic applications. The slightly less critical Guitar source signal is combined with the simulated (R)BRIRs from MUSHRA-REF and presented to a mixed group of listeners of the categories A, B and C. The experiment parameters are listed in Table 5.9. The results are depicted in Figure 5.13.

Table 5.9 Parameters MUSHRA-BCS.

Experiment ID	MUSHRA-BCS (Best-case scenario)
Test paradigm	MUSHRA
Varied parameter(s)	Plane waves (including reflections) with limited modal orders $N = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 17, 23\}$ for two different composite grids
Spatial filter(s)	Simulated RBRIRs
Reference	Simulated BRIRs
Anchor	Unfiltered diotic (including reflections)
Number of ratings	26
Listener grading	Mixed C(14), B(7), A(5)
Trials	1
No. of stimuli per trial	13 + Anchor + Hidden Reference
Source signal	Drums

The results show a rapid convergence towards the reference with increasing modal order. Even comparatively low orders achieve high ratings for less critical scenarios and

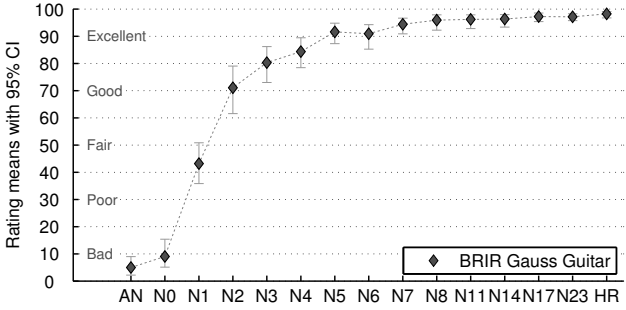


Figure 5.13 MUSHRA-BCS results.

a mixed group of listeners. The 90% (excellent) rating mark is already surpassed at order $N = 5$. Minor differences (<10%) apparently are audible up to much higher orders, even though the perceived differences at higher orders ($N \geq 7$) show no statistical significance when compared to the reference. The outcome of MUSHRA-BCS is quite positive for a practical construction of array systems for binaural auralization. Thus, microphone arrays with realistic expense and complexity can achieve satisfying perceptual results in this context, with the constraints of the limited modal resolution.

5.6.2 Radial Filter Limiting

Another important restriction is due to the excessive radial filter amplification demands that are discussed in Section 3.6.3. The required mode amplification for maintaining a constant array directivity quickly increases for higher orders at lower temporal frequencies. Since practical systems comprise equipment noise (tranducers, amplifiers, analog-digital converters), the radial filters need to be limited to rational amplification values in order to obtain a stable array signal. In Section 3.6.4 a soft-knee limiting approach is discussed and in Section 3.6.5 non-critical radial filters are proposed, which are depicted in Figure 5.14.

The mode amplification limit is set to $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$. Hence, arbitrary attenuation and no amplification is allowed for the critical range $f < (N\,c)/(2\,\pi\,r)$. This can be considered to be a very strict criterion. In return, this configuration ensures a definitely stable array response. Limiting the radial filter amplification leads to reduced spatial resolution of the array at lower temporal frequencies. If the array radius undercuts a certain limit, which was previously found to be in the range of $r_0 = 15\,\mathrm{cm}$ for non-critical radial filters, the HRTFs start to be successively truncated in their modal order in the affected frequency range. For a detailed discussion return to Section 3.6.8, where the impact of radial filter limiting on binaural reproduction is discussed.

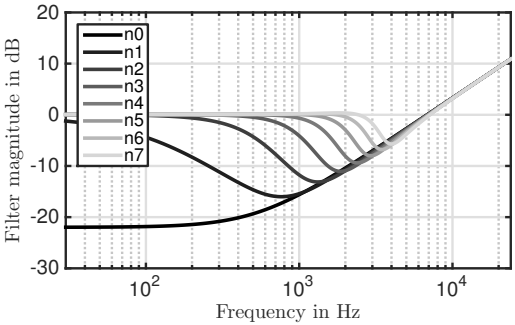


Figure 5.14 Magnitude response of the proposed non-critical radial filters. The filters are shifted along the frequency axis with varying array radius r_0.

The following experiments MUSHRA-RFL and SAQI-RFL aim at assessing the perceptual influences of applying non-critical radial filters to a system for array-based binaural reproduction.

5.6.2.1 MUSHRA-RFL

In MUSHRA-RFL, non-critical radial filters as depicted in Figure 5.14 are used for generating the stimuli. The radius of the simulated array is varied in the range of $r_0 = 2\,\text{cm}$ to $r_0 = 30\,\text{cm}$ in order to evaluate the perceptual influences of using a non-critical filter configuration in dependence of the array measurement radius. The reference is generated with unlimited[4] radial filters, which is only feasible for simulated data sets. The experiment parameters are listed in Table 5.10. The results are depicted in Figures 5.15 to 5.18.

The results of MUSHRA-RFL show the minimum microphone array radius that should be used for array based binaural auralization. Accordingly, the array should at least be of the order of the head radius or slightly above for a non-critical filter configuration with $\hat{a}_{\text{dB}} = 0\,\text{dB}$. A radius of $r_0 = 9\,\text{cm}$ achieves ratings around the 90% (excellent) mark. Smaller radii entail considerable artifacts and larger radii improve the signal only slightly. For radii $r_0 \geq 15\,\text{cm}$ none of the ratings show statistically significant differences compared to the reference. Thus, a reasonable array dimension for binaural auralization is in the range of $r_0 = 9\,\text{cm}$ to $r_0 = 15\,\text{cm}$, which exactly corresponds to our theoretical predictions. There is no significant difference between trials with different orders N.

[4] The limit is actually set to 250 dB in order to avoid numerical instabilities.

Table 5.10 Parameters MUSHRA-RFL.

Experiment ID	MUSHRA-RFL (Radial filter limiting)
Test paradigm	MUSHRA
Varied parameter(s)	Varying array radius $r_0 = 2 - 30\,\text{cm}$ with fixed radial filter limit of 0 dB for modal orders $N = \{4, 5, 6, 7\}$
Spatial filter(s)	RHRIRs with limited filters
Reference	RHRIRs with unlimited filters
Anchor	Diotic with hi-shelf filter (-12 dB, 2 kHz)
Number of ratings	21
Listener grading	Mixed B(15), A(6)
Trials	4 (one per modal order)
No. of stimuli per trial	13 + Anchor + Hidden Reference
Source signal	Drums

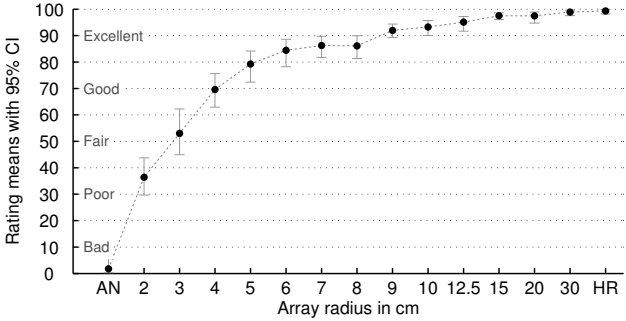


Figure 5.15 MUSHRA-RFL results for $N = 4$.

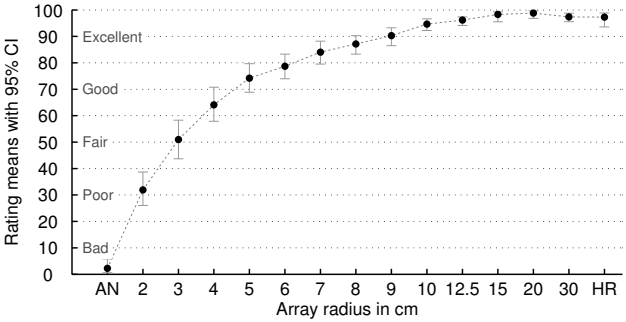


Figure 5.16 MUSHRA-RFL results for $N = 5$.

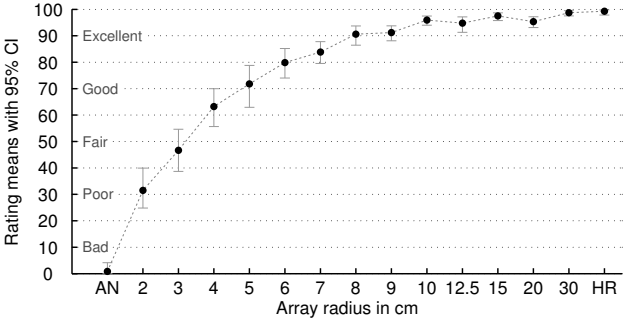


Figure 5.17 MUSHRA-RFL results for $N = 6$.

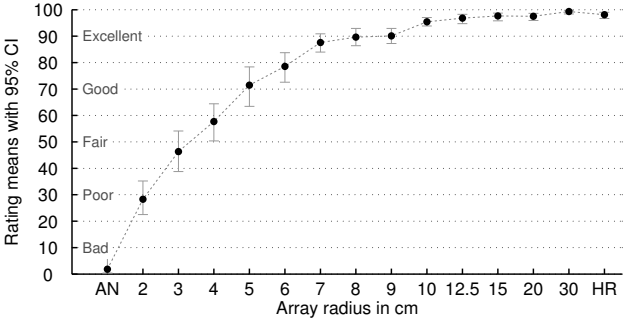


Figure 5.18 MUSHRA-RFL results for $N = 7$.

5.6.2.2 SAQI-RFL

SAQI-RFL is conducted for the purpose of a more specific examination of the perceptual artifacts that come along with radial filter limiting. The reference is a RHRTF of $N = 7$ with Gauss composite grid and unlimited radial filters. The stimulus is a RHRTF of $N = 7$ with Gauss composite grid and a radius of $r_0 = 5\,\mathrm{cm}$ with non-critical radial filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$. The identical stimulus achieved ratings of about 70% (good) in the MUSHRA-RFL experiment. Thus, the audio quality is still acceptable, but the stimulus comprises clearly audible artifacts. The experiment parameters are listed in Table 5.11. The results are depicted in Figure 5.19. They indicate coloration of the stimulus, i.e. decreased high/mid and increased low frequency tone color. Furthermore, a decrease of externalization and an additional horizontal shift become apparent. The source width is slightly decreased and a minimal dynamic compression effect is perceived.

Table 5.11 Parameters SAQI-RFL.

Experiment ID	SAQI-RFL (Radial filter limiting)
Test paradigm	SAQI
Varied parameter(s)	Array radius $r_0 = 5\,\mathrm{cm}$ with the radial filter limit set to $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ (non-critical)
Spatial filter(s)	RHRIRs ($N = 7$) with limited radial filters
Reference	RHRIRs ($N = 7$) with unlimited radial filters
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	Drums

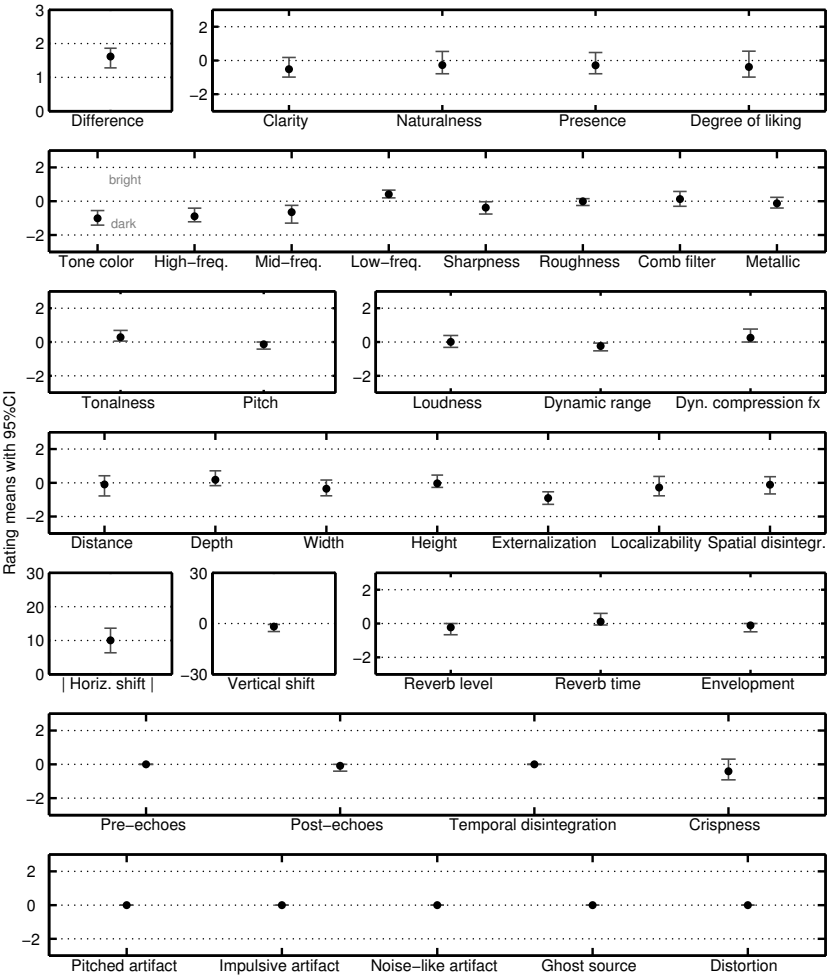


Figure 5.19 SAQI-RFL results.

5.6.3 Spatial Aliasing

Spatial aliasing emerges due to discrete spatial sampling of the sound field, which was discussed in Section 3.8. Depending on the respective spatial sampling density, higher spherical harmonic modes cannot be appropriately resolved. The resulting modal ambiguity manifests itself as spatial aliasing. Primarily higher temporal frequencies are affected by this, since they tend to excite increasingly higher modes. Spatial aliasing decreases or even impedes the spatial resolution of the microphone array within the affected temporal frequency bands. Different listening experiments were conducted in order to evaluate the perceptual influences of spatial aliasing.

5.6.3.1 MUSHRA-SA I

In MUSHRA-SA I, discrete spatial sampling is simulated in order to obtain simulated array responses including spatial aliasing artifacts. Even though spatial aliasing is usually present in the full temporal frequency spectrum, and, as a consequence, no strict aliasing-boundary can be defined, the aliasing contributions suddenly increase excessively above a certain aliasing frequency f_A which is approximately determined by Eq. (3.44). This equation indicates the dependency of f_A on the grid order N_sg and the array radius r_0. Increasing the grid order or decreasing the radius shifts the aliasing frequency f_A towards higher temporal frequencies. MUSHRA-SA I was conducted in order to evaluate the magnitude of perceptual artifacts coming from spatial aliasing. Two different stimulus traces were generated for testing the influences of the grid order N_sg and the radius r_0. Different grid orders $N_\mathrm{sg} = \{7, 11, 14, 17, 20\}$ of a Lebedev sampling grid were applied, using a fixed array radius of $r_0 = 200\,\mathrm{mm}$. Then, the grid order $N_\mathrm{sg} = 7$ is kept constant and the radius $r_0 = \{200, 127, 100, 82, 70\}\,\mathrm{mm}$ is varied according to Eq. (3.44) in order to obtain identical aliasing frequencies $f_\mathrm{A} = \{1.9, 3.0, 3.8, 4.6, 5.4\}\,\mathrm{kHz}$ for both traces. The start configuration is $N_\mathrm{sg} = 7$ and $r_0 = 200\,\mathrm{mm}$. The experiment parameters are listed in Table 5.12 and the results are depicted in Figure 5.20.

The results clearly prove that spatial aliasing is audible and they indicate that a proportional relation exists between the strength of perceptual artifacts and f_A. Furthermore, less perceptual artifacts are noticed when increasing the grid order (depending on the grid density) rather than decreasing the radius. According to Eq. (3.44) both should be equivalent for a constant target aliasing frequency f_A. In addition to the reduced spatial selectivity of the array by itself, an increasing spectral coloration for frequencies $f > f_\mathrm{A}$ arises, which can be deduced from the simulations in Section 3.8.2, as well as from the SAQI-SA experiment in Section 5.6.3.4.

Table 5.12 Parameters MUSHRA-SA I.

Experiment ID	MUSHRA-SA I (Spatial Aliasing)
Test paradigm	MUSHRA
Varied parameter(s)	Fixed array radius $r_0 = 20\,\text{cm}$ with varying sampling grid density/order $N_{\text{sg}} = \{7, 11, 14, 17, 20\}$ vs. varying radius $r_0 = \{20, 12.7, 10, 8.2, 7\}\,\text{cm}$ at a fixed grid order $N_{\text{sg}} = 7$ for common theoretical aliasing frequencies.
Spatial filter(s)	RHRIRs, simulating discrete spatial sampling
Reference	RHRIRs, ideal / no spatial sampling
Anchor	Diotic with hi-shelf filter (+15 dB, 2 kHz)
Number of ratings	21
Listener grading	Mixed B(15), A(6)
Trials	1
No. of stimuli per trial	9 + Anchor + Hidden Reference
Source signal	Drums

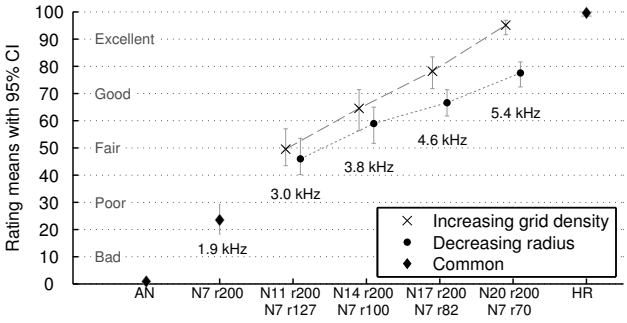


Figure 5.20 MUSHRA-SA I.

5.6.3.2 MUSHRA-SA II

In MUSHRA-SA II the influence of the sample grid's node distribution is examined. Two different node distributions, to a Gauss and a Levedev quadrature, are applied to simulating spatial sampling. The aliasing frequency f_A is kept constant at 2 kHz throughout the experiment by adaption of the radius r_0 according to Eq. (3.44). Different grid orders $N_\mathrm{sg} = \{4, 5, 6, 7\}$ were explored in separate trials. The experiment parameters are listed in Table 5.13 and the results are depicted in Figure 5.21.

Table 5.13 Parameters MUSHRA-SA II.

Experiment ID	MUSHRA-SA II (Spatial Aliasing)
Test paradigm	MUSHRA
Varied parameter(s)	Fixed aliasing frequency $f_\mathrm{A} = 2\,\mathrm{kHz}$ for two different sampling grids Gauss (Gss) and Lebedev (Lev) at RHRIR orders $N_\mathrm{sg} = \{4, 5, 6, 7\}$.
Spatial filter(s)	RHRIRs, simulating discrete spatial sampling
Reference	RHRIRs, ideal / no spatial sampling
Anchor	Diotic with hi-shelf filter (+15 dB, 2 kHz)
Number of ratings	21
Listener grading	Mixed B(15), A(6)
Trials	4
No. of stimuli per trial	2 + Anchor + Hidden Reference
Source signal	Drums

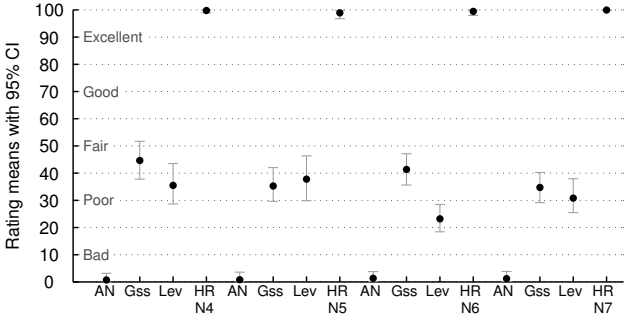


Figure 5.21 MUSHRA-SA II.

Except for $N_{\mathrm{sg}} = 6$, the ratings for different sampling grids do not show statistically significant differences. Thus, different node distributions of the spatial sampling grid apparently introduce a comparable amount and comparable characteristics of perceptual aliasing artifacts, as long as an identical aliasing frequency is established for both grids. This result stands in contrast to the outcome of the MUSHRA-MR experiments, where different node distributions of the composite grid entailed clearly significant perceptual differences. However, the spatial node distribution of the array sensors and the inherent subsampling artifacts appear to be nearly irrelevant compared to the influence of the node distribution of the composite grid.

5.6.3.3 MUSHRA-SA III

In the third experiment of this series, MUSHRA-SA III, the influence of additional early reflections on a spatially subsampled system are examined. Due to the loss of spatial selectivity in the aliased frequency bands, additional reflections could be expected to mask the spatial aliasing artifacts. The stimuli of MUSHRA-SA III are based on the simulated (R)BRIRs previously described in Section 5.6.1.5. The experiment parameters are listed in Table 5.14 and the results are depicted in Figure 5.22.

Table 5.14 Parameters MUSHRA-SA III.

Experiment ID	MUSHRA-SA III (Spatial Aliasing)
Test paradigm	MUSHRA
Varied parameter(s)	Varying radius $r_0 = \{20, 15, 10, 5\}$ cm with Lebedev sampling grid of order $N_{\mathrm{sg}} = 7$ using RHRIRs and RBRIRs of $N = 7$ with a Gauss composite grid.
Spatial filter(s)	RHRIRs/RBRIRs, simulating discrete spatial sampling
Reference	RHRIRs/RBRIRs, ideal / no spatial sampling
Anchor	Diotic with hi-shelf filter (+15 dB, 2 kHz)
Number of ratings	21
Listener grading	Mixed B(15), A(6)
Trials	2
No. of stimuli per trial	4 + Anchor + Hidden Reference
Source signal	Drums

The ratings of the RBRIRs are slightly higher than those of the RHRIRs, but the differences between the rating means are not statistically significant. Hence, the inherent spatial aliasing artifacts are perceived nearly equal in a dry virtual source and in a virtual source with additional early reflections. Again, much of the perceivability of the

spatial aliasing artifacts can be ascribed to the predominant coloration of the source signal. This is affirmed in the SAQI-SA test in the next section. When appropriately compensating for the coloration, and adding reflections and diffuse reverberation to a virtual source, spatial aliasing in higher temporal frequency bands entails a surprisingly small perceptual impairment only, refer to the experiments in Section 5.7.

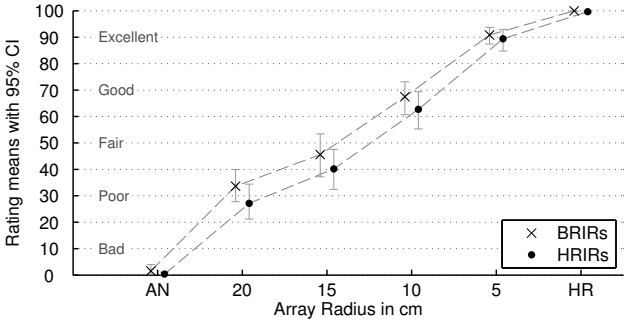


Figure 5.22 MUSHRA-SA III.

5.6.3.4 SAQI-SA

A RHRTF set of order $N = 7$ with a Gauss composite grid and with a Gauss sampling grid $N_{\mathrm{sg}} = 7$ on a radius of $r_0 = 10\,\mathrm{cm}$ for simulating spatial sampling was presented to the participants. The aliasing frequency is located around $f_{\mathrm{A}} = 3.8\,\mathrm{kHz}$ for this configuration. The same stimulus achieves ratings of around 60% (between fair and good) in the MUSHRA-SA I test. Thus, spatial aliasing artifacts are clearly perceivable but the signal is still acceptable. The equivalent RHRTF set without simulating spatial sampling and without spatial aliasing artifacts is presented as reference. The experiment parameters are listed in Table 5.15 and the results are depicted in Figure 5.23.

Table 5.15 Parameters SAQI-SA.

Experiment ID	SAQI-SA (Spatial Aliasing)
Test paradigm	SAQI
Varied parameter(s)	RHRIR set with $N = 7$, Gauss composite grid and Gauss sampling grid of $N_{\mathrm{sg}} = 7$ on a radius $r_0 = 10\,\mathrm{cm}$ simulating discrete spatial subsampling vs. an equivalent RHRIR set without simulating spatial sampling
Spatial filter(s)	RHRIRs including spatial subsampling ($f_{\mathrm{A}} = 3.8\,\mathrm{kHz}$)
Reference	RHRIRs without spatial sampling
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	Drums

The overall perceived difference is large. The results confirm a strong coloration of the signal in the higher frequency range. This reflects in a bright, sharp and metallic tone color. The source is perceived closer to the subject, less externalized, and spatially disintegrated. Primarily the width, but also the depth and height of the source are increased and the localizability is decreased. A minor increment of the source elevation is perceived. The stimulus appears louder than the reference. Pitched artifacts, noise-like artifacts, and increased distortion are perceived. The signal appears to be slightly pitched. Minor post-echoes and temporal disintegration of the source are detected.The aliasing artifacts are apparently interpreted as a minor increase in reverberation, even though neither of the presented stimuli involved reverberation.

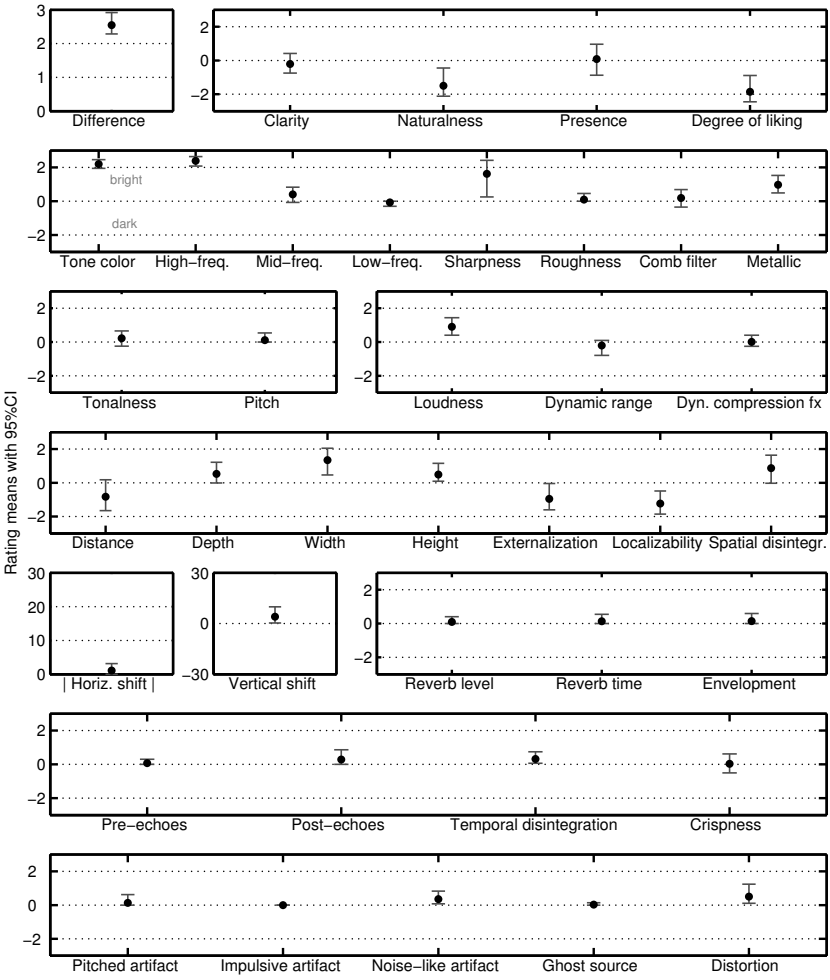


Figure 5.23 SAQI-SA results.

5.6.4 MUSHRA-BEMA (Anti-Aliasing)

In the MUSHRA-BEMA experiment, the performance of the BEMA approach of Section 3.11 for patching aliased frequency bands in simulated signals is evaluated. The BEMA anti-aliasing method is discussed in Section 3.11. For a single isolated plane wave incidence the BEMA method allows for an exact reconstruction of the aliased frequency bands. Hence, a listening test is obsolete in this case. For multiple plane waves that arrive from arbitrary directions and with arbitrary phase relation at once, the array signal can only be approximated and not reconstructed exactly. Thus, a virtual source with additional reflections according to Section 5.6.1.5 is used for evaluating the performance of the BEMA method.

Table 5.16 Parameters MUSHRA-BEMA.

Experiment ID	MUSHRA-BEMA (Spatial anti-aliasing)
Test paradigm	MUSHRA
Varied parameter(s)	Varying radius $r_0 = \{20, 15, 10, 5\}$ cm with Lebedev sampling grids of orders $N_{\mathrm{sg}} = \{5, 7\}$ using RBRIRs with Gauss composite grids. Raw stimuli with spatial aliasing vs. stimuli with BEMA anti-aliasing.
Spatial filter(s)	RBRIRs, simulating discrete spatial sampling with and without BEMA anti-aliasing
Reference	RBRIRs, ideal / no spatial sampling
Anchor	Diotic with hi-shelf filter (+15 dB, 2 kHz)
Number of ratings	21
Listener grading	Mixed B(15), A(6)
Trials	2
No. of stimuli per trial	4 + Anchor + Hidden Reference
Source signal	Drums

A temporal block size of 128 samples with zero samples overlap is used for partitioning the simulated array input signals in the time domain. The internal BEMA processing core upscaled the blocks to a FFT size of 8192 samples for extracting an appropriately resolved spatio-temporal image down to low temporal frequencies. The extraction bandwidth is set to 1/3 octave around the most stable array frequency range at $f = (N_{\mathrm{sg}}\, c)/(2\,\pi\, r)$. In the MUSHRA-BEMA experiment, two trials with different sampling grid orders $N_{\mathrm{sg}} = \{5, 7\}$ were performed. For each trial, the radius of the simulated array is varied in four steps, $r_0 = \{20, 15, 10, 5\}$ cm. The resulting array configurations entailed different aliasing frequencies $f_{A,N5} = \{1.4, 1.8, 2.7, 5.5\}$ kHz and $f_{A,N7} =$

$\{1.9, 2.5, 3.8, 7.6\}$ kHz, respectively. Based on the configuration, a stimulus pair is generated that consists of an aliased raw stimulus and an equivalent stimulus that is treated with the BEMA anti-aliasing procedure for comparison. The experiment parameters are listed in Table 5.16 and the results are depicted in Figure 5.24 ($N_{\mathrm{sg}} = 5$) and Figure 5.25 ($N_{\mathrm{sg}} = 7$).

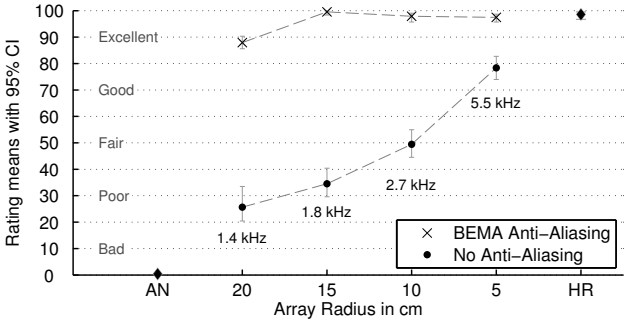


Figure 5.24 MUSHRA-BEMA, $N_{\mathrm{sg}} = 5$.

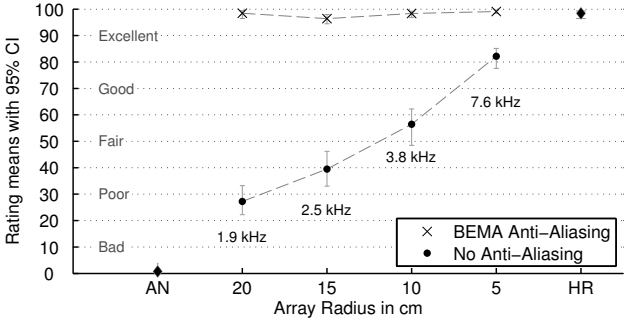


Figure 5.25 MUSHRA-BEMA, $N_{\mathrm{sg}} = 7$.

The results show a highly significant improvement of the stimuli with applied BEMA anti-aliasing processing in comparison to the aliased raw stimuli. None of the stimuli using BEMA anti-aliasing indicates statistically significant perceptual differences to the reference, with exception of $N_{\mathrm{sg}} = 5$ and $r_0 = 20\,\mathrm{cm}$. The BEMA patching seems to work well at both orders $N_{\mathrm{sg}} = \{5, 7\}$. Apparently, the processing can be applied at relatively low aliasing frequencies, like $f_{\mathrm{A}} \approx 2\,\mathrm{kHz}$. The low rating for $N_{\mathrm{sg}} = 5$ and $r_0 = 20\,\mathrm{cm}$ could either be a random audible artifact for the specific configuration, or indicate a lower valid boundary of f_{A} for using BEMA. Further experiments need to be

conducted in order to answer that question. Generally, the results are quite promising and proof the basic concept of the BEMA procedure, at least for simulated signals of comparably low complexity.

5.7 Measured Data Sets

In this section several listening experiments based on measured microphone array and dummy head data sets are presented. The data sets are introduced in Section 4.3.3. Two rooms with different acoustic properties were chosen for the experiments. The first room is a large broadcast studio (LBS) with a volume of around 6500 m^3 and reverberation times around 1.7 s. The second room is a studio control room (CR7) with a room volume of around 170 m^3 and reverberation times around 0.25 s. A portable directional sound system[5] was used for excitation in the broadcast studio.

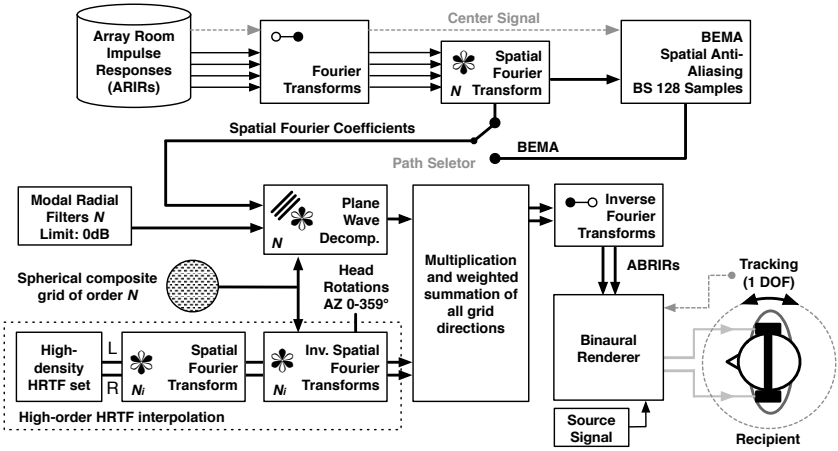


Figure 5.26 System architecture for processing the measured data sets and presenting the stimuli.

In the control room, a speaker of the main monitoring system[6] was used instead. The capturing device is a VariSphear scanning microphone array system, refer to Section 4.1. The corresponding BRIR references were captured at a common pivot point with a Neumann KU 100 dummy head mounted on the VariSphear motion base. For further details concerning the locations and the technical setup refer to Section 4.3.3. The

[5] Name in the library (Stade et al., 2012): LBS-PAC
[6] Name in the library (Stade et al., 2012): CR7-L

system architecture and processing chain for generating the ABRIRs is described in Figure 5.26.

5.7.1 Array Configuration and Filters

The array was operated in rigid sphere configuration with a sphere diameter of $d_0 = 17.5\,\mathrm{cm}$. This diameter was used for capturing the full audio bandwidth. Non-critical radial filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$, introduced in Section 3.6.5, are applied. The configuration is not optimized. According to the considerations and outcomes from Section 3.6.8 and the listening experiments in Section 5.6.2, the radius is too small for binaural auralization. The radial filter limit could have been raised to compensate for the small radius. But a larger measurement radius was not available due to constructional constraints. Non-critical filters with $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ are used in order to have a reasonable reference amplification. The array did not form an ideal sphere, as the motor and rotation mount stick out of the rigid sphere body, which is discussed in Section 3.13.6. A spectral compensation filter, refer to Figure 3.63, is applied to the output signal in order to compensate for the impaired frequency response. In this light, the ratings from the following listening experiments can be expected to be slightly lower than the ones that were achievable with an optimized array with an ideal sphere of appropriate size or using less strict radial filter amplification limits.

Full resolution ABRIRs were generated based on a Lebedev sampling grid of order $N_{\mathrm{sg}} = 29$, comprising 1202 spatial sampling nodes. By theory, according to Eq. (3.44), this configuration is stable up to $f_{\mathrm{A}} = 18\,\mathrm{kHz}$. Simulations prove the constellation to be stable for frequencies up to at least $f_{\mathrm{A}} = 20\,\mathrm{kHz}$ with negligible aliasing contributions. Hence, no considerable spatial aliasing arises on the entire audible frequency range for this configuration. The plane wave decomposition operations are performed on maximum orders $N = \{4, 5, 6, 7\}$, depending on the respective scenario. Order-adapted Gauss quadratures are used as composite grid for all experiments, since they achieved best perceptual ratings in the MUSHRA-MR experiments of Section 5.6.1.

5.7.2 Motion-tracked Binaural Sound (MTB)

Stimuli based on the motion tracked binaural sound (MTB) method were added to the experiments for comparison. The MTB method is introduced in Section 1.1.7.

In order to minimize the influences of the number of microphones and interpolation methods, a virtual high density MTB array with 360 sampling nodes lined up on the equator was used for generating the stimuli. This yielded a resolution of 1° in the horizontal plane. Since this is identical to the resolution of the reference BRIR sets, no additional MTB-specific interpolation needed to be performed. The MTB signals are

derived using high-order spherical harmonic interpolation of high-density array data sets captured with the VariSphear rigid-sphere head at a diameter of $d_0 = 17.5\,\text{cm}$, which corresponds to the average head diameter.

In order to minimize spectral coloration of the MTB signals compared to the original BRIRs, specific global spectral compensation filters were generated based on magnitude averaging of the circular sets. Applying the compensation filters, the MTB set comprises the same spectral circular diffuse field response as the BRIR reference set and the perceived difference in global coloration vanishes. MTB stimuli with spectral compensation filters are marked as «MTB EQ» in the plots, whereas raw MTB stimuli without compensation filters are denoted as «MTB». No further customization or individualization of the MTB stimuli as proposed in (Melick et al., 2004) is performed. No torso reflections are present, which is not relevant in the given context, as the Neumann KU 100 reference dummy head does not provide a torso either.

Considering the very high microphone density to avoid cross-fading or interpolation artifacts, and the precise spectral equalization to the reference BRIR set, the presented MTB signals can be regarded as idealized MTB stimuli. Most of the remaining differences can be ascribed mainly to the missing pinna cues and other «anatomic» differences between the reference dummy head and the MTB array.

5.7.3 Measured ABRIRs vs. BRIRs

In the first listening experiment, MUSHRA-MMR, ABRIR sets with varied modal orders are compared to the corresponding reference BRIR sets. The array configuration and filter settings for generating the ABRIRs are given in Section 5.7.1. Since the sampling grid order is sufficiently high ($N_{\text{sg}} = 29$, 1202 nodes), spatial aliasing can be considered negligible in the entire audible frequency spectrum. The experiments were conducted in a broadcast studio and in a studio control room. For comparison of the results, compensated MTB EQ sets, cf. Section 5.7.2, are included with each trial.

5.7.3.1 MUSHRA-MMR I

In MUSHRA-MMR I, ABRIR sets for the broadcast studio and the control room were explored as described above. The experiment parameters are listed in Table 5.17. The results for both rooms are depicted in Figure 5.27.

The ratings for both rooms are very similar and no statistically significant differences can be seen. Generally, the differences between ABRIRs and the reference BRIRs are statistically significant. The rating means for $N = 7$ surpass the 90% (excellent) mark, which indicates perceptible but very small differences. Further analysis of the perceptual properties concerning this stimulus are presented in Section 5.7.3.3. The results of

Table 5.17 Parameters MUSHRA-MMR I

Experiment ID	MUSHRA-MMR I. (Modal reduction based on measured data)
Test paradigm	MUSHRA
Varied parameter(s)	Varying modal order $N = \{4, 5, 6, 7\}$ of ABRIRs based on measured high-resolution microphone array data from two acoustically distinct rooms
Spatial filter(s)	ABRIRs, MTB
Reference	BRIRs
Anchor	Diotic
Number of ratings	24
Listener grading	Mixed B(17), A(7)
Trials	2 (1 trial per room)
No. of stimuli per trial	5 + Anchor + Hidden Reference
Source signal	*Drums*

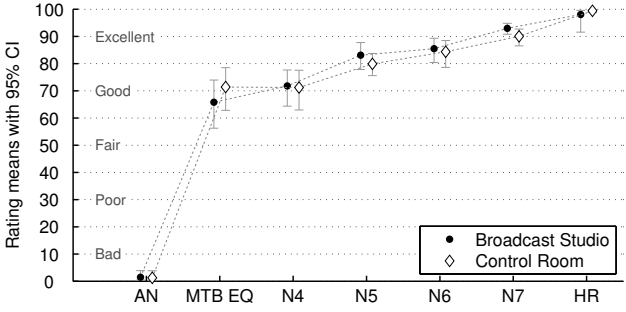


Figure 5.27 MUSHRA-MMR I.

MUSHRA-MMR I are comparable and consistent with the outcomes of the MUSHRA-MR experiments in Section 5.6.1, which are based on simulated array responses. The ratings are proportional to the order of the ABRIRs, i.e. higher modal orders achieve higher ratings. Even the absolute range of the ratings is well-comparable to the previous experiments that are based on simulated data sets. This points to a certain reliability and validity of the listening experiments, the signal processing chain, and the measured data sets. The ratings for the optimized spectrum-compensated MTB EQ stimulus are located on about the same level as the ratings for ABRIRs with $N = 4$. The ABRIRs for orders $N \geq 5$ achieve significantly improved perceptual results when compared to the MTB array signals.

5.7.3.2 MUSHRA-MMR II

In the MUSHRA-MMR II experiment, the data sets from the broadcast studio are split into two temporal sections. The first segment (a) includes the direct sound and some early reflections (ERs). The second segment (b) involves further ERs and the full diffuse reverberation tail, cf. Figure 5.28.

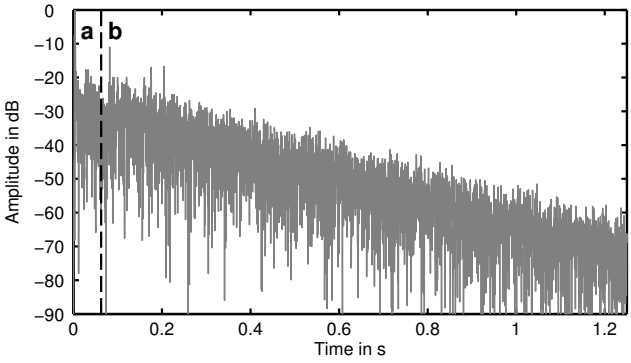


Figure 5.28 Energy-time curve (16x downsampled) from the broadcast studio, using a directional source and an omni-directional receiver located at the reference origin. Segment **a** includes the direct sound and early reflections from the floor and the side walls. Segment **b** includes dedicated back plane and ceiling reflections, as well as the entire diffuse reverberation tail. The transition is at 62 ms.

The experiment should give information on the influences of either dedicated singular incident waves (i.e. especially the direct sound), on the one hand, and largely diffuse sound fields, on the other hand. The experiment parameters are listed in Table 5.18 and the results are depicted in Figure 5.29. The graph shows combined ratings for

Table 5.18 Parameters MUSHRA-MMR II.

Experiment ID	MUSHRA-MMR II (Modal reduction based on measured data)
Test paradigm	MUSHRA
Varied parameter(s)	Varying modal order $N = \{4, 5, 6, 7\}$ of ABRIRs based on measured high-resolution microphone array data from the broadcast studio. The sets are split at approx. 62 ms into two sections, cf. Figure 5.28
Spatial filter(s)	ABRIRs, MTB (temporal segments)
Reference	BRIRs (temporal segments)
Anchor	Diotic
Number of ratings	24
Listener grading	Mixed B(17), A(7)
Trials	2 (1 per section) + 1 from MUSHRA-MMR I
No. of stimuli per trial	5 + Anchor + Hidden Reference
Source signal	*Drums*

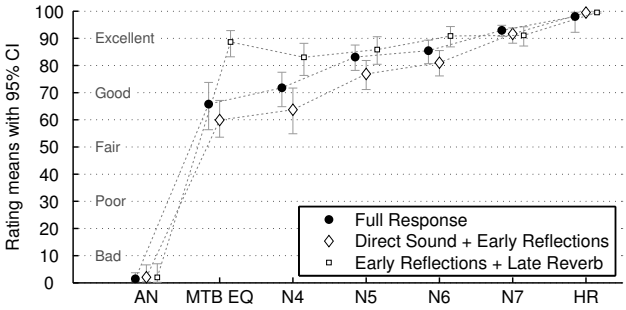


Figure 5.29 MUSHRA-MMR II (Broadcast Studio).

both trials performed in segment a and segment b. Additionally, the related ratings for MUSHRA-MMR I are plotted for reference, thus providing the ratings for the full response.

The results of MUSHRA-MMR II show decreased ratings for segment a (direct sound) and increased ratings for segment b (reverberation) compared to the values from MUSHRA-MMR I, where the full response is presented. The perceived artifacts of the ABRIRs can apparently be primarily ascribed to the direct sound. Even though the ratings for segment b (reverberation) are generally higher and appear more constant, a certain remaining dependence on the order can be observed. Generally, all of the reverberation stimuli from segment b show statistically significant differences to the reference. This can be ascribed to the strong back plane and ceiling reflections in segment b, cf. Figure 5.28, and to possible remaining spectral differences even after applying the spectral compensation filters. The filters were globally generated for the full response and not specifically adapted to the single segments a or b. Since the ratings for the full response are mostly located in between the two segment ratings, the reverberation can be assumed to partly mask the perceived artifacts on the direct sound, which entails increased overall ratings for the full response. The ratings in segment b (reverberation) for the MTB EQ stimulus are comparably high, which corresponds to expectations, as the missing monaural cues are of minor importance for the perception of diffuse sound fields. Nevertheless, the MTB EQ ratings of the direct sound, as well as of the full response are clearly located below the ratings of any ABRIR with $N \geq 5$. Both segment ratings for the ABRIR with $N = 7$ surpass the 90% (excellent) mark, indicating perceivable but very small differences.

5.7.3.3 SAQI-MMR

The ABRIR with $N = 7$ turns out to be perceived quite similar to the original reference. However, minor differences remain. The SAQI-MMR experiment is conducted in order to gain a deeper insight into the remaining differences. An ABRIR with $N = 7$ from the broadcast studio is compared to the reference BRIR set for this purpose. The experiment parameters are listed in Table 5.19 and the results are depicted in Figure 5.30.

The results of SAQI-MMR confirm rather small perceptual differences between the ABRIRs and the reference HRIRs. Even the category-A listeners reported difficulties in assigning definite differences. This is reflected in the distribution of the rating values. Most of the ratings show very small displacements randomly scattered around the zero level. Besides indicating typical uncertainties that arise for very similar stimuli, the distribution around the zero level could possibly be ascribed to the dynamic behavior

Table 5.19 Parameters SAQI-MMR.

Experiment ID	SAQI-MMR (Modal reduction based on measured data)
Test paradigm	SAQI
Varied parameter(s)	Measured ABRIR set with $N = 7$, Gauss composite grid and Lebedev sampling grid of $N_{\mathrm{sg}} = 29$ on a diameter $d_0 = 17.5\,\mathrm{cm}$ vs. measured reference BRIR set.
Spatial filter(s)	ABRIR
Reference	BRIR
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	*Drums*

of the ABRIRs depending on the wave incidence direction or the listener's rotational angle.

Good absolute agreement between all listeners can be seen for a minimal amount of temporal disintegration, arising pre/post echoes, as well as for the manifestation of a ghost source. This can be ascribed to the artifacts of the modal order reduction. The source generally appears to be shifted in the horizontal plane, which could primarily be ascribed to the applied radial filter limiting, when considering the outcomes of SAQI-RFL in Section 5.6.2.1. A minor decrease in crispness is perceived. All in all, the differences between both stimuli appear to be small and difficult to grasp. Without a direct A/B comparison to an original reference BRIR set, a decisive detection of the ABRIR set can be supposed to be even more difficult. All in all, an ABRIR set with $N = 7$ can hardly be distinguished from a corresponding original BRIR set, even by critical listeners under good listening conditions.

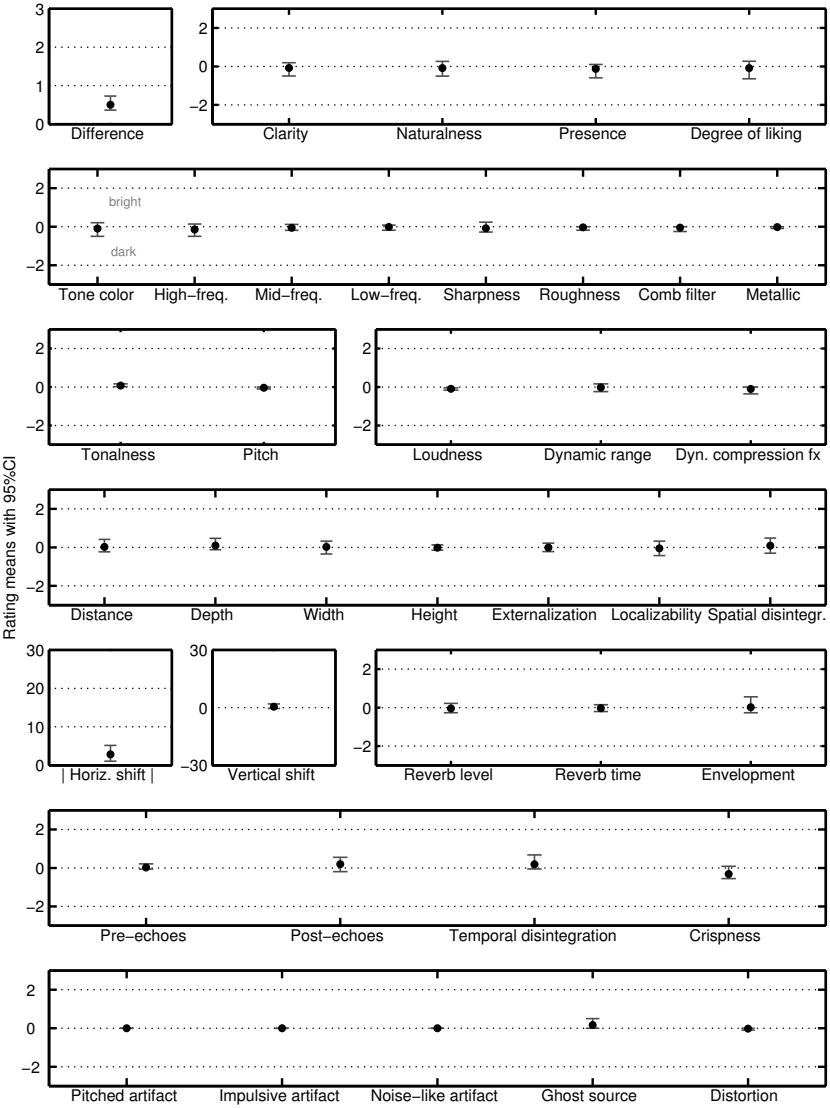


Figure 5.30 SAQI-MMR results.

5.7.4 ABRIRs with Reduced Sensor Density

The previous experiments are based on high-density impulse response data sets that involve as many as 1202 spatial sampling nodes. Hence, spatial aliasing contributions are negligible in the entire audible spectrum for the given sphere diameter of $d_0 = 17.5\,\text{cm}$. The data sets were acquired with a scanning array system that allows for capturing arbitrary sampling density at no extra cost but time. However, such sensor density cannot be realized with real-time arrays in practice. As long as no alternate technical solution for sampling the sound field is found, the system must work at a reduced sensor density.

Table 5.20 Legend MUSHRA-RSD stimuli.

Stimulus name	Explanation
AN	Anchor
MTB	Motion tracked binaural sound
MTB EQ	Motion tracked binaural sound with spectral compensation
A86	Aliased stimulus based on 86 sampling nodes
A86 EQ	Aliased stimulus based on 86 sampling nodes with spectral compensation filters
A110	Aliased stimulus based on 110 sampling nodes
A110 EQ	Aliased stimulus based on 110 sampling nodes with spectral compensation
BEMA	Aliased stimulus based on 86 sampling nodes with BEMA processing applied to the aliased frequency band
1202	Stimulus based on 1202 sampling nodes (no spatial aliasing)
HR	Hidden reference

The minimum sensor density is determined by the desired target order for generating the ABRIRs, which is assumed to be $N = 7$ in the following. The minimum radius is determined by the radial filters. According to the MUSHRA-RFL experiment in Section 5.6.2.1, the array radius should not be much below $r_0 = 10\,\text{cm}$. The aliasing frequency of such a configuration is typically located around $f_\text{A} = 4\,\text{kHz}$, which is clearly located in the audible spectrum. The MUSHRA-SA and SAQI-SA experiments in Section 5.6.3 indicate considerable impairments for the aliased stimuli; for measured signals with diffuse reverberation and appropriate spectral compensation fewer impairments can be expected.

For the following experiments, two subsampled Lebedev sampling grids with 86 nodes ($N_\text{sg} = 7$) and 110 nodes ($N_\text{sg} = 8$), respectively, are used. The corresponding aliasing

frequencies for a sphere diameter of $d_0 = 17.5\,\text{cm}$ are $f_\text{A} = 4.4\,\text{kHz}$ and $f_\text{A} = 5\,\text{kHz}$, respectively. Since larger contributions to the perceived impairments for spatial aliasing originate from spectral coloration, cf. Section 3.8, the aliased stimuli are presented in two versions, i.e. with and without spectral compensation filters, cf. Section 3.8.2.

The BEMA anti-aliasing procedure from Section 3.11.2 is applied to the stimuli based on the data sets with 86 nodes. The algorithm is set to 1/3 octave extraction bandwidth at around $f_\text{A} \approx 4.4\,\text{kHz}$ and using a block size of 128 samples with zero samples overlap. The blocks are internally upscaled to 8192 samples to increment the resolution at low frequencies. The required omni-directional center signal was captured at the physical origin of the array in a separate measurement.

MTB stimuli with and without spectral compensation are used according to Section 1.1.7. The names of the stimuli in the plots are explained in Table 5.20.

5.7.4.1 MUSHRA-RSD I

In MUSHRA-RSD I, stimuli from the broadcast studio and the control room were presented in two subsequent trials. The experiment parameters are listed in Table 5.21 and the results are depicted in Figure 5.31.

Table 5.21 Parameters MUSHRA-RSD I.

Experiment ID	MUSHRA-RSD I (Reduced sensor density)
Test paradigm	MUSHRA
Varied parameter(s)	Number of spatial sampling nodes, compensation filters, BEMA anti-aliasing processing
Spatial filter(s)	ABRIRs, MTB
Reference	BRIRs
Anchor	Diotic with hi-shelf filter (+15 dB, 4 kHz)
Number of ratings	24
Listener grading	Mixed B(17), A(7)
Trials	2 (Broadcast studio and control room)
No. of stimuli per trial	8 + Anchor + Hidden Reference
Source signal	*Drums*

As expected from the MUSHRA-SA and SAQI-SA experiments in Section 5.6.3, both aliased stimuli, A86 and A110, achieve ratings between 50-60% (fair to good) only. The A110 ratings are located slightly above the A86 ratings. As soon as spectral compensation filters are applied (A86 EQ and A110 EQ), the ratings improve significantly

and achieve values larger than 80% (good to excellent). This confirms the prevalence of spectral coloration artifacts due to spatial aliasing. There appears to be no perceivable difference between using 86 or 110 sampling nodes. Apparently, even microphone arrays with a considerably reduced sensor density can achieve good perceptual quality for binaural auralization, as long as an appropriate filter for spectral compensation of spatial aliasing artifacts is applied. A slight increase of the number of sensors does not entail any advantage from a perceptual point of view. Hence, using the minimum required sensor density for achieving the target order is a reasonable choice. Both facts are very relevant for the construction of microphone arrays for binaural auralization.

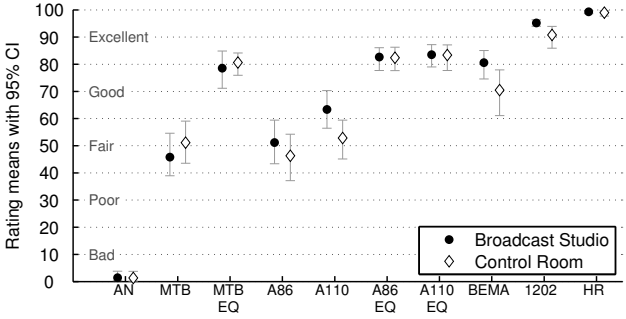


Figure 5.31 MUSHRA-RSD I.

The BEMA ratings are higher than the A86 and A110 ratings, but stay below the ratings for A86 EQ and A110 EQ. This is surprising in view of the excellent ratings of the MUSHRA-BEMA experiments from Section 5.6.4. A closer inspection reveals the presence of audible processing artifacts of the BEMA algorithm applied to measured signals This is confirmed in the SAQI-BEMA experiment in Section 5.7.4.4. The processing artifacts can be ascribed to the block segmentation. They become particularly audible for complex or diffuse portions of the response. This explains the discrepancy between the ratings of MUSHRA-RSD I and the MUSHRA-BEMA experiment in Section 5.6.4, since the MUSHRA-BEMA experiment does not involve diffuse reverberation. The BEMA artifacts depend on the parametrization of the algorithm, i.e. the extraction bandwidth, the block sizes, and the overlap. A consistent parametrization, which in informal pretests delivered reasonable results, was chosen for all experiments. In summary, the BEMA processing further on appears to be a promising approach; the implementation needs to be refined in order to suppress the audible segmentation artifacts that cause the rather moderate ratings.

The MTB stimuli achieve ratings around 50% (fair) without spectrum adaptation and 80% (good to excellent) with spectrum adaptation. The spectrum adaptation is important for removing global spectral differences compared to the reference BRIRs, which is reflected in the ratings. The MTB approach delivers results comparable to the aliased ABRIRs. This puts in question the motivation to fall back to the considerably more elaborate procedure using spherical arrays (with reduced sensor density). However, the ABRIRs have several advantages.

The MTB stimuli are generated with as many as 360 spatial sampling nodes, which needs to be reduced in practice. Hence, the disadvantages of cross-fading or interpolation algorithms (Lindau and Roos, 2010) arise, which are expected to entail additional impairments for the MTB stimuli. Also, the ABRIRs are not restricted to the azimuthal plane, and make possible full three-dimensional head rotation including elevation. By contrast, the MTB signals are restricted to the azimuthal plane in this case. Thirdly, the ABRIRs allow for the use of individual HRTFs with little effort; individual HRTFs need to be captured for the few nodes of the composite grid only. Only 128 individual HRTF nodes need to be acquired with a Gauss composite grid of order $N_{\mathrm{sg}} = 7$ in order to obtain an individual array system for binaural auralization that enables full three-dimensional head rotation. Hence, spherical arrays and sound file decomposition techniques for binaural auralization are worth the increased effort compared to MTB.

In summary, falling back to a system with a certain amount of spatial aliasing and spectral compensation filters is a reasonable choice for binaural auralization in practice. The impairments due to spatial aliasing appear to be tolerable, as long as the aliasing frequency is sufficiently high. Further approaches, such as spatial anti-aliasing filters as discussed in Section 3.8.1 and Section 3.10.2, could potentially improve the perceptual results.

5.7.4.2 MUSHRA-RSD II

In MUSHRA-RSD II, the stimuli from the broadcast studio are divided into two sections as shown in Figure 5.28, analogous to the MUSHRA-MMR II experiment in Section 5.7.3.2. The first section involves the direct sound and early reflections and the second section involves further early reflections and late diffuse reverberation. The experiment parameters are listed in Table 5.22 and the results are depicted in Figure 5.32.

The results demonstrate the perceived artifacts that are introduced by the BEMA processing, especially in diffuse reverberation. The ratings are significantly lower than the ratings for the aliased and spectrum-compensated stimuli A86 EQ and A110 EQ, as well as for the MTB EQ stimulus. The latter achieve very good ratings >90% (excellent) for the reverberation and acceptable ratings between 70% and 85% (good

to excellent) for the direct sound. The direct sound of the A110 stimulus is rated slightly higher than the direct sound of the A86 stimulus.

Table 5.22 Parameters MUSHRA-RSD II.

Experiment ID	MUSHRA-RSD II (Reduced sensor density)
Test paradigm	MUSHRA
Varied parameter(s)	Number of spatial sampling nodes, compensation filters, BEMA anti-aliasing processing
Spatial filter(s)	ABRIRs, MTB (temporal segments)
Reference	BRIRs (temporal segments)
Anchor	Diotic with hi-shelf filter (+15 dB, 4 kHz)
Number of ratings	24
Listener grading	Mixed B(17), A(7)
Trials	2 (1 per section) + 1 from MUSHRA-RSD I
No. of stimuli per trial	8 + Anchor + Hidden Reference
Source signal	*Drums*

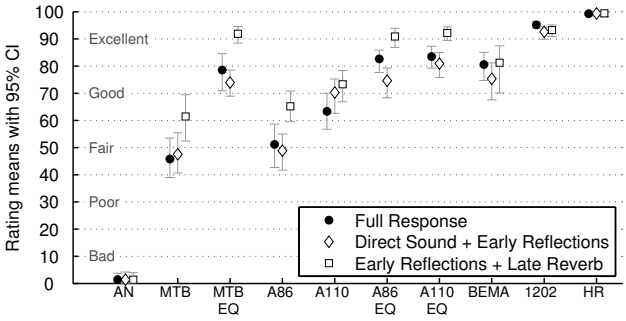


Figure 5.32 MUSHRA-RSD II.

5.7.4.3 SAQI-AEQ

The A86 EQ stimuli achieve acceptable ratings in the MUSHRA experiments. They are particularly interesting for the construction of systems in practice, as they are based on a realistic number of microphones. The SAQI-AEQ experiment is conducted in order to gain a better understanding of the perceptual artifacts.

The A86 EQ set is compared to the reference BRIR set in the broadcast studio. The experiment parameters are listed in Table 5.23 and the results are depicted in Figure 5.33.

The overall differences are well-perceivable. The results indicate minor coloration of the stimulus; the mid frequency range appears to be slightly more emphasized and the low frequency range to be reduced. Minimal manual equalization beyond the analytic spectrum compensation filters could remove the residual coloration. The stimulus is perceived brighter, with more sharpness and a minor additional comb-filter artifact. The dynamic range seems to be slightly reduced, even though no dynamic processing is applied to the signal. The source is perceived closer to the listener and the externalization decreases slightly. Temporal disintegration of the source and the manifestation of a ghost source are perceived. The signal appears to be less crisp and the source is shifted in the horizontal plane. The naturalness decreases. All in all, the aliased signal is clearly different from the reference, but the impairment appears to be perceived as hardly annoying, as the participants do not dislike the stimulus at all.

Table 5.23 Parameters SAQI-AEQ.

Experiment ID	SAQI-AEQ (Spatial aliasing with spectrum compensation)
Test paradigm	SAQI
Varied parameter(s)	Measured ABRIR set with $N = 7$, Gauss composite grid and Lebedev sampling grid of $N_{\mathrm{sg}} = 7$ on a diameter $d_0 = 17.5\,\mathrm{cm}$ vs. measured reference BRIR set.
Spatial filter(s)	ABRIR
Reference	BRIR
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	*Drums*

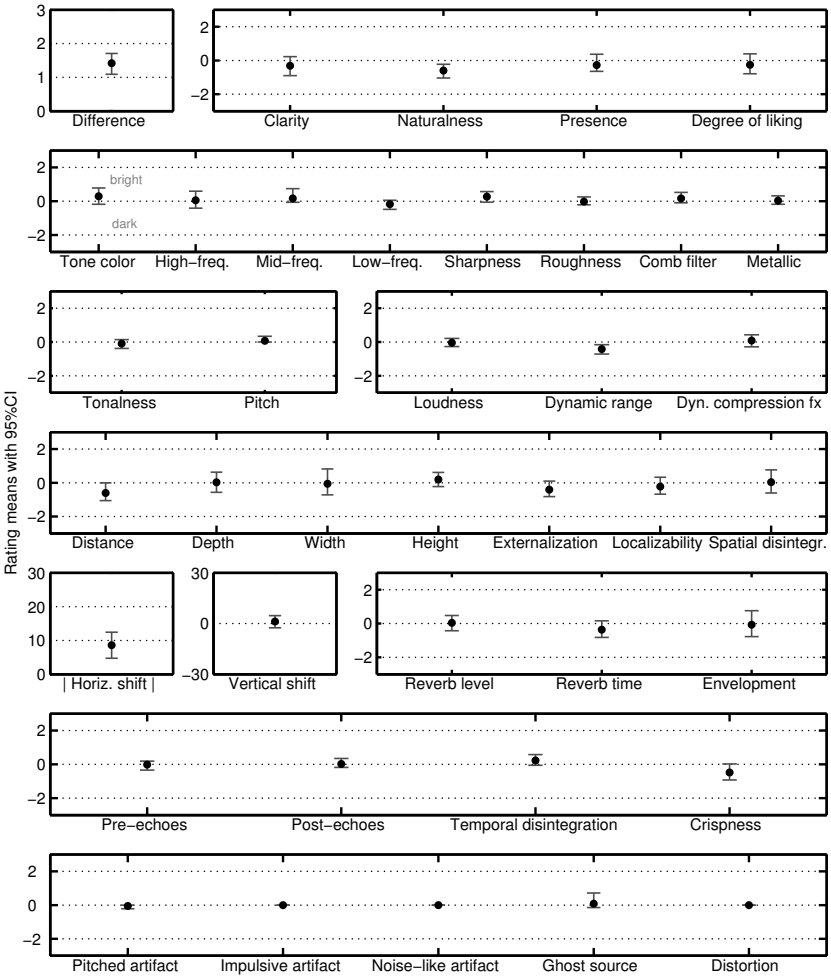


Figure 5.33 SAQI-AEQ results.

5.7.4.4 SAQI-BEMA

The BEMA processing appears to introduce audible processing artifacts to the signal. The SAQI-BEMA experiment is conducted in order to gain a deeper insight into the processing artifacts. The BEMA-treated ABRIR set from MUSHRA-RSD I in Section 5.7.4.1 is compared to the BRIR reference set from the broadcast studio.

Table 5.24 Parameters SAQI-BEMA.

Experiment ID	SAQI-BEMA (Spatial aliasing with BEMA anti-aliasing)
Test paradigm	SAQI
Varied parameter(s)	Measured ABRIR set with $N = 7$, Gauss composite grid and Lebedev sampling grid of $N_{\mathrm{sg}} = 7$ on a diameter $d_0 = 17.5\,\mathrm{cm}$ with BEAM anti-aliasing above approx 4 kHz. vs. measured reference BRIR set.
Spatial filter(s)	ABRIR
Reference	BRIR
Number of ratings	11
Listener grading	A
Trials	1
No. of stimuli per trial	1 + Reference
Source signal	*Drums*

The experiment parameters are listed in Table 5.24 and the results are depicted in Figure 5.34. The results show considerable impairments of the stimulus. All global attributes are clearly decreased. A slight reduction of high frequencies and a darker tone color is perceived. The processing artifacts manifest themselves in increased roughness, comb-filter sound, metallic sound color, post echoes, temporal disintegration, a slight distortion, and additionally perceived impulsive artifacts. The dynamic range appears to be decreased with audible dynamic compression artifacts.

The source dimensions tendentially decrease and the source appears to be smaller, which is quite remarkable. Also, the source is less externalized. The reverberation is clearly affected by the BEMA processing, since the reverberation level, time, and envelopment decrease. Additionally, a horizontal shift of the source, decreased crispness, and the manifestation of a ghost source are perceived. However, the latter attributes supposedly can be ascribed to the modal reduction rather than to the BEMA processing. All in all, the BEMA anti-aliasing approach entails considerable artifacts, at

least using a proper but quite basic implementation of the algorithm. Further development and refinement of the algorithm needs to be conducted in order to get rid of the undesired processing artifacts.

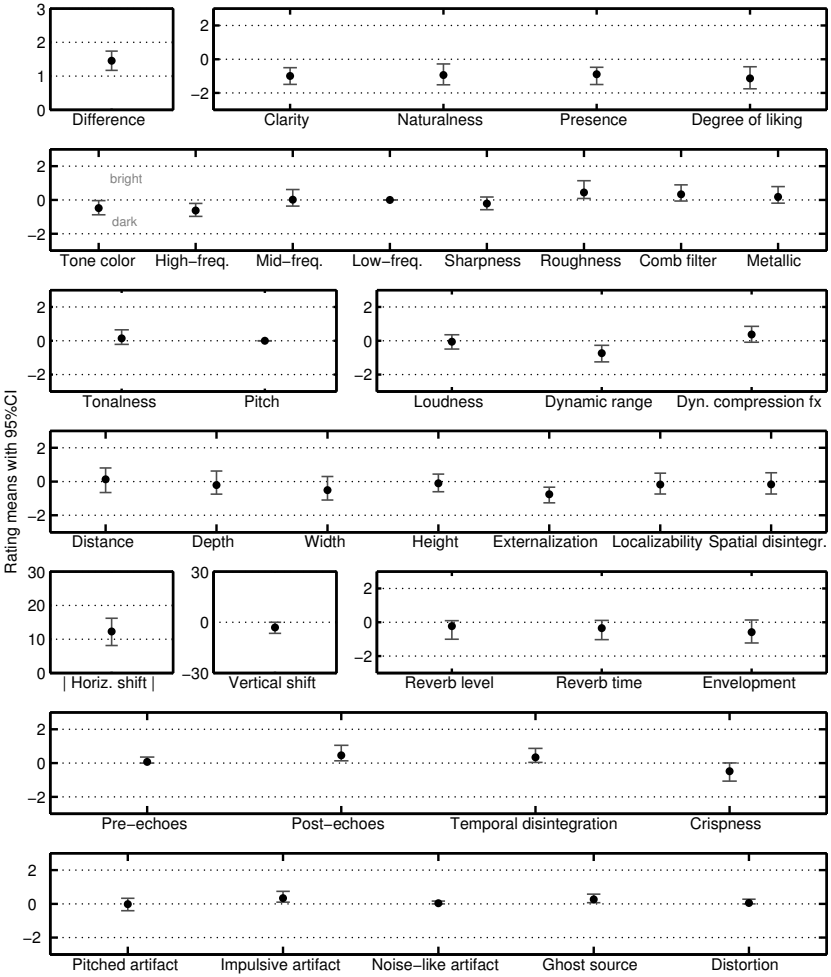


Figure 5.34 SAQI-BEMA results.

6 Summary and Conclusions

In the following, essential approaches, results and, conclusions are summarized.

6.1 Theory

A theoretical transparent closed-form solution for deriving ideal binaural signals from a sound field description defined on a spherical surface S_0 does exist. The approach is based on spherical harmonic decomposition of the sound field and HRTF-weighted recombination. The method allows for factoring in individual HRTFs, as well as performing free horizontal, vertical, and lateral head rotation in an independent post-processing stage. In theory, even position transitions of the listener in the sound field can be realized.

6.2 Constraints in Technical Systems

According to the current state of technology, spherical microphone arrays can be used for sampling the sound field on the surface S_0 in practice. Array-based binaural systems can be realized that extent the capabilities of the classical dummy head by allowing for head rotation and by using individual HRTFs for multiple recipients. However, there are constraints inherent in technical systems, in particular in microphone arrays, which impede maintaining ideal conditions as in theory.

Microphone arrays require discrete spatial sampling. As a consequence, we obtain sound field descriptions with limited modal resolution. It is shown that sound field descriptions with lower modal resolution than the expansion of the HRTF set are basically incompatible with binaural reproduction. An expansion of a HRTF set requires approximately spherical harmonic orders of about $N = 35$, while realistic microphone arrays typically achieve orders of $N = [4, 7]$ only. The inherent order-truncation leads to low-pass effects in the time-frequency domain. It is shown that subsampling the HRTF set using a composite grid of the highest order included in the sound field description considerably decreases order-truncation artifacts and improves the overall transmission properties.

Another consequence of discrete spatial sampling is spatial aliasing. Spatial aliasing manifests itself as additive spatial noise being present over the entire time-frequency range. The contributions increase in proportion to the temporal frequency. We can

locate an aliasing frequency f_A that depends on the specific properties of the sampling grid and on the array radius, where aliasing contributions start to surge excessively. Below f_A spatial aliasing can be neglected, whereas above f_A the signal is massively impaired. The spatial selectivity of the array is lost and increased output levels can be observed in this range.

Different approaches to overcome spatial aliasing or at least to improve the transmission properties in the presence of aliasing are discussed. The trivial solutions consist of increasing the sensor density or decreasing the radius. Increasing the sensor density is expensive, since the required number of sensors increases in quadratic relation with the temporal frequency. Decreasing the radius is only feasible within certain limits that are determined by noise constraints and the modal intensity distribution of HRTFs. Hence, with a realistic array radius and a realistic number of sensors, we must expect serious aliasing contributions from around 4 kHz to 5 kHz upwards.

Two different methods to improve the transmission properties are proposed in this thesis. The first one is to accept the spatial impairments and to equalize the diffuse field response in order to minimize the coloration that is inherent to spatial aliasing. The second approach is a method entitled bandwidth extension for microphone arrays (BEMA) and is based on the separate acquisition and treatment of spatiotemporal and spectrotemporal properties of the sound field. The spatiotemporal information is extracted from a particularly robust section of the array response, whereas the spectrotemporal information is acquired using a separate omni-directional transducer. New Fourier coefficients are derived for patching impaired sections of the response, which are fully compatible with the original coefficients. The approach has several requirements and restrictions, and a refined implementation would be required to avoid audible block-processing artifacts in diffuse sound fields.

Microphone arrays deliver signals with finite signal-to-noise ratio (SNR). As a consequence, the radial filter amplification must be limited in order to achieve a robust array response. This entails successively decreased modal resolution towards lower temporal frequencies. Theoretical considerations, as well as comprehensive analysis of the white noise gain (WNG) show the specific impact of uncorrelated noise in the transducer paths on array-based binaural auralization. We discovered that the analysis of the WNG of single plane wave decomposition (PWD) signals is not conclusive, since the array works in composite mode, i.e. it provides a closed-form recombination of the decomposed PWD signals. This aspect needs to be considered for most kind of error analyses. A general analytic expression for predicting the WNG of array-based binaural systems is derived.

It is shown that the reduced modal resolution due to limiting the radial filter amplification does not affect the binaural signal, as long as we provide a suitable combination of array radius and radial filter amplification limit that lies within realistic dimensions. As a final conclusion we notice that an array for binaural auralization should provide a radius of $r_0 \approx 10\,\text{cm}$. This specific radius should not be markedly undercut in order not to truncate the modal expansion of HRTFs in the mid-frequency band. This radius should also not be markedly exceeded either, since this increases noise at the top frequency end and entails more spatial aliasing contributions. As a rule of thumb, the optimum size for an array for binaural auralization is comparable to the size of a human head, which appears quite plausible.

Spatial aliasing and limited radial filter amplification yield a limited effective operational bandwidth (EOB) of about 0.5 oct to 1 oct. Different ways of increasing the EOB, such as using more sensors, raising the radial filter amplification limit, or employing multiple array spheres of different diameters, are discussed. Illustration show the dependence of the EOB on different influencing factors. Even though for classical array applications, like beamforming, the achievable EOB of a single-sphere array appears to be rather poor, we see that it might be sufficient for performing reasonable array-based binaural auralization.

An expanded transducer that deviates from the infinitesimal point transducer, assumed in theory, produces a sinc-like weighting of the modes in the spherical wave spectrum domain. If the transducer expansion is large in relation to the measurement radius r_0, complete dropouts of the signal power arise for dedicated modes owing to the zeros in the sinc-function. However, for array dimensions like $r_0 = 10\,\text{cm}$ and typical 1/8" or 1/4" microphone capsules, the impact of the capsule expansion turns out to be negligible for the considered orders. From a different point of view, expanded transducers can be applied to create basic modal low-pass filters in order to reduce spatial aliasing. Passive sub-arrays of multiple joint transducers covering a certain surface could be applied to every single node in order to realize simple modal low-pass filters.

There are additional influencing factors, such as sources in the near-field, non-ideal transducers, positioning inaccuracies, time variances in scanning arrays, incomplete sampling, or non-ideal sphere configurations.

6.3 Technology and Resources

Since the approaches are relatively recent, not yet established in commercial products, and the subject of current research, there was hardly any suitable ready-made hard- or software available at the time of writing this thesis. Therefore, a spherical microphone

array measurement system (VariSphear) was designed and built, a sound field analysis toolbox (SOFiA) was designed, implemented, and verified, and appropriate data sets were acquired and validated. A suitable environment for listening experiments was set up. The implemented software and the acquired data sets were made accessible to the scientific community under either GNU GPL or Creative Commons licenses.

6.4 Listening Experiments

Several listening experiments were performed in order to evaluate the performance of the approach and to assess the implications of certain technical constraints on perception.

A broad series of tests is dedicated to the binaural reproduction of sound field descriptions with limited modal order. The test results are not only of particular interest for array-based binaural auralization, but conclusive for binaural auralization of any modal sound field description, such as higher-order Ambisonics (HOA) desciptions.

The tests indicate significant perceptual improvements at low wave orders when using the proposed HRTF subsampling compared to high-order composite grids for adaptation or truncating the modal HRTF order, which is equivalent. HRTF subsampling is applied to all experiments yielding either head-related transfer functions with limited modal resolution (RHRTFs) or binaural room impulse responses with limited modal resolution (RBRIRs).

The exact structure and node distribution of the composite grid influences the perceptual properties at (singular) lower decomposition orders. Two different composite grid types are analyzed, an equiangular Gauss quadrature and an equidistant Lebedev quadrature. While the Gauss composite grid shows good, predictable and stable performance, the Lebedev grid shows volatile performance that varies with the specific decomposition order. The Lebedev composite grid does not show a significantly improved performance compared to the Gauss grid for none of the stimuli. Hence, the Gauss composite grid is preferable for array-based binaural auralization in all cases.

The perceived similarity of RHRTF and HRTF quickly rises for orders between $N = 0$ and $N = 11$. From approximately $N = 11$ upwards only a minor tendential difference remains, which, however, is not statistically significant. Hence, a RHRTF or RBRIR with $N \geq 11$ can be considered to be identical to the original full-resolution counterpart from a perceptual point of view. The remaining perceived differences appear to be minimal and can only be detected in ideal surroundings and critical scenarios.

At low decomposition orders, the order-reduction artifacts in RHRTFs and RBRIRs appear to be well-perceivable and yield more or less comparable ratings, no matter

whether only direct sound, direct sound and early reflections, or a full response including direct sound, early reflections and diffuse reverberation is presented. Simulated and measured sound fields have similar ratings in general.

Decomposition orders $N < 5$ yield substantial perceptual impairment and can be regarded as improper for binaural auralization. Orders starting from $N = 5$ appear to be acceptable but not fully satisfying. The most reasonable trade-off between technical effort and perceptual properties might be found at $N = 7$. The required number of microphones for realizing a real-time array of that order is well within realistic technical and economic limits. The MUSHRA ratings for $N = 7$ are very close to, or even above, the 90%-excellent mark in all scenarios. Some perceivable differences remain in a direct A-B comparison, but the achieved results appear to be satisfying. A SAQI (Lindau et al., 2014) test is employed to assess more specific differences between an original BRIR set and a measured array-based RBRIR set of $N = 7$. Even critical expert listeners perceived only subtle differences. We conclude that $N = 7$ would actually be a reasonable target order for designing realistic array-based binaural systems that achieve satisfying perceptual results.

Limiting the radial filter amplification is indispensable in order to achieve a robust array response in practice. Non-critical radial filters are proposed, with a strict limit of $\hat{a}_{\mathrm{dB}} = 0\,\mathrm{dB}$ for $\omega \leq (N\,c)/r_0$ and full native amplification in the range of $\omega > (N\,c)/r_0$. In theory, we assume that the non-critical radial filters do not yield considerable impairments of the binaural signal, as long as the array provides a minimum radius $r_0 \geq 15\,\mathrm{cm}$.

The theoretical assumptions are confirmed by the results of listening experiments. With non-critical radial filters no significant differences can be perceived when using an array radius of $r_0 \geq 15\,\mathrm{cm}$. For $r_0 = 9\,\mathrm{cm}$ the mean ratings are located around the 90%-excellent mark. For smaller radii the signal is impaired and the ratings decrease quickly. With non-critical radial filters, $r_0 = 9\,\mathrm{cm}$ can be regarded as the minimum array radius for binaural auralization. There are perceivable impairments, but the signal still appears to be satisfying.

The non-critical radial filters are excessively restrained. The analysis of the white noise gain (WNG) indicates a certain headroom for raising the radial filter amplification limit while still preserving reasonable WNG at low frequencies. This slightly increases the effective operational bandwidth (EOB) and allows for decreasing the radius. Exhausting the headroom of reasonable radial filter gain, we conclude that an radius of approximately $r_0 = 10\,\mathrm{cm}$, or even $r_0 = 9\,\mathrm{cm}$, does not yield perceivable impairments in the binaural output signal.

Spatial aliasing is perceivable as soon as the aliasing frequency f_A is located in the audible range. The structure of the spatial sampling grid apparently does not have much influence on the perceptual performance when the aliasing frequency is kept constant. There are two major aspects concerning the impact of spatial aliasing. The first and obvious aspect is the loss of spatial selectivity in the aliased range, which leads to perception of increased source width or spatial disintegration. The second aspect is an increased output level in the aliased range. The listening experiments show the perception of spectral coloration. Even if this effect is less obvious than the loss of spatial selectivity, it turns out to be the predominating factor from a perceptual point of view.

In order to counter the spectral coloration, the use of a simple global diffuse field compensation filter is proposed. The listening experiments show significant improvements with a spectral compensation filter. As soon as the coloration is removed, even aliased stimuli yield good to excellent ratings. Even the effects of widened sources or spatial disintegration appear to decrease considerably. The spatial aliasing frequency should no be too low, in order not to affect vital localization cues. However, even admitting a certain amount of spatial aliasing at higher frequencies (e.g. $f > 4\,\mathrm{kHz}$), reasonable binaural auralization with more or less satisfying perceptual properties can be performed while just using appropriate global spectral compensation filters.

Using the BEMA approach for patching aliased sections of the array response yields highly significant perceptual improvement for simulated sound fields without diffuse components. While untreated stimuli show significant perceptual differences to the reference, BEMA-treated stimuli are not distinguishable from the reference in most cases. This proves the validity of the BEMA approach. However, perceivable block processing artifacts arise in measured sound fields that provide diffuse components. As a consequence, the BEMA-treated stimuli finally not achieve improved ratings compared to aliased stimuli with spectral compensation filters. To the contrary, SAQI tests show that aliased stimuli with spectral compensation filters yield more pleasant perceptual results than BEMA-treated stimuli. Further refinement of the BEMA implementation could possibly reduce the perceived artifacts.

The stimuli based on sound field decomposition without spatial aliasing ($M = 1202$ sensors) with $N = 7$ yield significantly higher ratings than the MTB stimuli with $M = 360$ sensors. The stimuli based on sound field decomposition including spatial aliasing ($M = 86$ or $M = 110$ sensors) with spectral aliasing compensation using $N = 7$ achieve comparable ratings with the MTB stimulus. The sound field decomposition approach provides a description that inherently allows for performing full horizontal, vertical and lateral head rotation without increasing the number of sensors, while the MTB

approach is limited to the horizontal plane if no additional sensors are used. In contrast to the MTB method, the sound field decomposition method allows for using individual HRTFs in a true closed-form approach. We conclude that sound field decomposition has several advantages over MTB. There is greater flexibility and the perceptual properties are better or at least equal when allowing a certain amount of spatial aliasing. The only disadvantage compared to MTB is the increased computational demand for the spatial Fourier transforms and the plane wave decomposition operations.

6.5 Applications

With the presented methods, the use of the classical classical dummy head is extended by head-tracking abilities and by individual HRTFs, which widely removes the known problems of classical binaural recording. The methods allow for live transmission as well as recording and delayed playback of a sound scene. The signals can be directly embedded and transmitted using available formats and codecs, such as MPEG-H Audio (Herre et al., 2014). They are adaptable to the available transmission bandwidth by a floating reduction or increment of the number of spatial Fourier coefficients, using a varying modal order. This is of great practical relevance for the use of the Internet as a broadcast channel.

As outlined in the prolog at the very beginning of this work, there are different practical applications for this technology, such as binaural point-to-multipoint transmission of live concerts, sport events, or theater plays, as well as advanced 3D-audio teleconferencing. Suitable microphone arrays can be built, according to the specifications developed in this thesis. Current commercial CPUs provide sufficient processing power to implement the required encoders and decoders. Spherical binaural decoders can be implemented in smart phones and the headphones could be equipped with head-tracking sensors in order to reach a broad range of users. The technology indeed could be commercially successful in a near future.

6.6 Final Conclusion

Systems for dynamic binaural recording based on microphone arrays and sound field decomposition techniques with satisfying perceptual properties can be realized within feasible technological and economical limits.

Bibliography

Abhayapala, T. D. (2008). Generalized Framework for Spherical Microphone Arrays: Spatial and Frequency Decomposition. *Proceedings of the ICASSP, IEEE*.

Abhayapala, T. D., Kennedy, R. A., and Williamson, R. C. (1999). Spatial Aliasing for Near-Field Sensor Arrays. *Electronics Letters, IET*, 35(10).

Abhayapala, T. D. and Ward, D. B. (2002). Theory and Design of High Order Sound Field Microphones Using Spherical Microphone Array. *Proceedings of the ICASSP, IEEE*.

Abramowitz, M. E. and Stegun, I. A. E. (1972). *Handbook of Mathematical Functions*. Courier Dover Publications.

AES (2015). AES69-2015: AES Standard For File Exchange - Spatial Acoustic Data File Format.

Agmon, M., Rafaely, B., and Tabrikian, J. (2009). Maximum Directivity Beamformer for Spherical-Aperture Microphones. *Proceedings of the WASPAA, IEEE*.

Ahrens, J. (2010). The Single-Layer Potential Approach Applied to Sound Field Synthesis Including Cases of Non-Enclosing Distributions of Secondary Sources. *PhD Thesis, Technical University of Berlin*.

Algazi, V. R., Avendano, C., and Duda, R. O. (2001a). Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies. *Journal of the Acoustical Society of America, JASA*, 109(3).

Algazi, V. R., Dalton, R. J., Duda, R. O., and Thompson, D. M. (2005). Motion-Tracked Binaural Sound for Personal Music Players. *Proceedings of the AES Convention*, 119, Paper 6557.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (2004). Motion-Tracked Binaural Sound. *Journal of the Audio Engineering Society*, 52(11).

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avenda, C. (2001b). The CIPIC HRTF Database. *Proceedings of the WASPAA, IEEE*.

Alon, D. L. and Rafaely, B. (2012). Spherical Microphone Array with Optimal Aliasing Cancellation. *Proceedings of the IEEEI Convention, IEEE*.

Arfken, G. B. and Weber, H. J. (2005). *Mathematical Methods for Physicists*. Elsevier Academic Press.

Atkinson, K. and Han, W. (2012). *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Springer.

Avni, A., Ahrens, J., Geier, M., Spors, S., Wierstorf, H., and Rafaely, B. (2013). Spatial Perception of Sound Fields Recorded by Spherical Microphone Arrays with Varying Spatial Resolution. *Journal of the Acoustical Society of America, JASA*, 133(5).

Balmages, I. and Rafaely, B. (2007). Open-Sphere Designs for Spherical Microphone Arrays. *Transactions on Audio, Speech, and Language Processing, IEEE*, 15(2).

Bauck, J. and Cooper, D. H. (1992). Generalized Transaural Stereo. *Proceedings of the AES Convention*, 93, Paper 3401.

Bauck, J. and Cooper, D. H. (1996). Generalized Transaural Stereo and Applications. *Journal of the Audio Engineering Society, JAES*, 44(9).

Baumgartner, R., Pomberger, H., and Frank, M. (2011). Practical Implementation of Radial Filters for Ambisonic Recordings. *Proceedings of the ICSA International Conference on Spatial Audio*.

Beerends, R. J., ter Morsche, H. G., van den Berg, J. C., and van den Vrie, E. M. (2003). *Fourier and Laplace Transforms*. Cambridge University Press.

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. Academic Press.

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. *Journal of the Audio Engineering Society, JAES*, 49(10).

Beranek, L. L. and Sleeper Jnr, H. P. (1946). The Design and Construction of Anechoic Sound Chambers. *Journal of the Acoustical Society of America, JASA*, 18(1).

Berkhout, A. J. (1988). A Holographic Approach to Acoustic Control. *Journal of the Audio Engineering Society, JAES*, 36(12).

Berkhout, A. J., de Vries, D., and Sonke, J. J. (1997). Array Technology for Acoustic Wave Field Analysis in Enclosures. *Journal of the Acoustical Society of America, JASA*, 102(5).

Bernschütz, B. (2012a). Bandwidth Extension for Microphone Arrays. *Proceedings of the AES Convention*, 133, Paper 8751.

Bernschütz, B. (2012b). Map Projections for the Graphical Representation of Spherical Measurement Data. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B. (2013a). A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B. (2013b). VariSphear - Quick Start Manual R13-0109.C. *Cologne University of Applied Sciences*.

Bernschütz, B. (2014). Adaptation of HRTFs to Plane Waves with Reduced Modal Order. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2010). Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und virtual Audio. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2011a). SOFiA Sound Field Analysis Toolbox. *Proceedings of the ICSA International Conference on Spatial Audio*.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2011b). Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2011c). Zeitinvarianzen durch Temperaturveränderung bei sequentiellen sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren. *Proceedings of the German DAGA Conference, DEGA*.

Bernschütz, B., Stade, P., and Rühl, M. (2012). Sound Field Analysis in Room Acoustics. *Proceedings of the VDT International Convention*.

Bernschütz, B., Vázquez-Giner, A., Pörschmann, C., and Arend, J. (2014). Binaural Reproduction of Plane Waves with Reduced Modal Order. *Acta Acustica united with Acustica*, 100(5).

Blackstock, D. T. (2000). *Fundamentals of Physical Acoustics*. Wiley Interscience.

Blauert, J. (1997). *Spatial Hearing*. The Psychophysics of Human Sound Localization. The MIT Press.

Blauert, J. (2013). *The Technology of Binaural Listening*. Springer.

Blauert, J., Braasch, J., Bucholz, J., Colburn, H. S., Jekosch, U., Kohlrausch, A., Mourjopoulos, J., Pulkki, V., and Raake, A. (2010). Aural Assessment by Means of Binaural Algorithms - The AABBA Project. *Proceedings of the ISAAR*, (2009).

Blauert, J. and Laws, P. (1978). Group Delay Distortions in Electroacoustical Systems. *Journal of the Acoustical Society of America, JASA*, 63(5).

Bortz, J. and Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- Und Sozialwissenschaftler*. Springer.

Bowman, J. J., Senior, T. B. A., Uslenghi, P. L. E., and Asvestas, J. S. (1970). Electromagnetic and Acoustic Scattering by Simple Shapes. North-Holland Publishing Company, Wiley Intercience.

Brinkmann, F., Lindau, A., Vrhovnik, M., and Weinzierl, S. (2014). Assessing the Authenticity of Individual Dynamic Binaural Synthesis. *Proceedings of the Joint Symposium on Auralization and Ambisonics, EAA*.

Burkhard, M. D. and Sachs, R. M. (1975). Anthropometric Manikin for Acoustic Research. *Journal of the Acoustical Society of America, JASA*, 58(1).

Chen, J., Van Veen, B. D., and Hecox, K. E. (1992). External Ear Transfer Function Modeling: A Beamforming Approach. *Journal of the Acoustical Society of America, JASA*, 92(4).

Ciba, S., Wlodarski, A., Herder, M., Rotter, A., Blickensdorff, J., Brinkmann, F., Vrhovnik, M., and Lindau, A. (2014). WhisPER User Documentation v1.8.0. *Technical University of Berlin*.

Ciba, S., Wlodarski, A., and Maempel, H.-J. (2009). WhisPER - A New Tool for Performing Listening Tests. *Proceedings of the AES Convention*, 126, Paper 7749.

Condon, E. U. and Shortley, G. H. (1951). *The Theory of Atomic Spectra*. Cambridge University Press.

Cooper, D. H. and Bauck, J. L. (1989). Prospects for Transaural Recording. *Journal of the Audio Engineering Society, JAES*, 37(1-2).

Damaske, P. (1971). Head-Related Two-Channel Stereophony with Loudspeaker Reproduction. *Journal of the Acoustical Society of America, JASA*, 50.

Daniel, J. (2003). Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format. *Proceedings of the AES Conference on Audio Recording and Reproduction*, 23, Paper 16.

Daniel, J., Nicol, R., and Moreau, S. (2003). Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. *Proceedings of the AES Convention*, 114, Paper 5788.

D'Appolito, J. (1999). *Lautsprecher-Meßtechnik*. PC-gestützte Analyse analoger Systeme. Elektor.

de Fornel, F. (2001). *Evanescent Waves: From Newtonian Optics to Atomic Optics*. Springer.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*, 11(3).

DIN (2009). DIN EN ISO 3382-1: Acoustics-Measurement of Room Acoustic Parameters-Part 1: Performance Spaces.

Du, Q., Gunzburger, M. D., and Ju, L. (2003). Constrained Centroidal Voronoi Tessellations for Surfaces. *SIAM Journal on Scientific Computing*, 24(5).

Duraiswami, R., Li, Z., Zotkin, D. N., Grassi, E., and Gumerov, N. A. (2005a). Plane-Wave Decomposition Analysis for Spherical Microphone Arrays. *Proceedings of the WASPAA, IEEE*.

Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (2004). Interpolation and Range Extrapolation of HRTFs. In *Proceedings of the ICASSP, IEEE*.

Duraiswami, R., Zotkin, D. N., Li, Z., Grassi, E., Gumerov, N. A., and Davis, L. S. (2005b). High Order Spatial Audio Capture and its Binaural Head-Tracked Playback over Headphones with HRTF Cues. *Proceedings of the AES Convention*, 119, Paper 6540.

Edmonds, A. R. (1957). *Angular Momentum in Quantum Mechanics*. Princeton University Press.

Elko, G. W. and Meyer, J. (2009). Second-Order Differential Adaptive Microphone Array. *Proceedings of the ICASSP, IEEE*.

Engdegard, J., Resch, B., Falch, C., Hellmuth, O., Hilpert, J., Hoelzer, A., Terentiev, L., Breebaart, J., Koppens, J., Schuijers, E., and Oomen, W. (2008). Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding. *Proceedings of the AES Convention*, 124, Paper 7377.

Evans, M. J., Angus, J. A., and Tew, A. I. (1997). Spherical Harmonic Spectra of Head-Related Transfer Functions. *Proceedings of the AES Convention*, 103, Paper 4571.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics*. Facts and Models. Springer.

Feynman, R. P., Leighton, R., and Sands, M. (2011). The Feynman Lectures on Physics. Basic Books.

Fisher, E. and Rafaely, B. (2008). The Nearfield Spherical Microphone Array. *Proceedings of the ICASSP, IEEE*.

Fisher, E. and Rafaely, B. (2009). Dolph-Chebyshev Radial Filter for the Near-Field Spherical Microphone Array. *Proceedings of the WASPAA, IEEE*.

Fisher, E. and Rafaely, B. (2011). Near-Field Spherical Microphone Array Processing with Radial Filtering. *Transactions on Audio, Speech, and Language Processing, IEEE*, 19(2).

Fliege, J. and Maier, U. (1999). The Distribution of Points on the Sphere and Corresponding Cubature Formulae. *IMA Journal of Numerical Analysis*, 19(2).

Fourier, J. B. J. (1822). *Théorie Analytique de la Chaleur (The Analytic Theory of Heat)*. Chez Firmin Didot, Father and Sons.

Gardner, W. G. (1997). 3-D Audio Using Loudspeakers. *PhD Thesis, Massachusetts Institute of Technology MIT*.

Geier, M., Ahrens, J., and Spors, S. (2008). The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. *Proceedings of the AES Convention*, 124, Paper 7330.

Geier, M., Ahrens, J., and Spors, S. (2010). Object-based Audio Reproduction and the Audio Scene Description Format. *Organised Sound*, 15(3).

Geier, M. and Spors, S. (2012). Spatial Audio with the SoundScape Renderer. *Proceedings of the VDT International Convention.*

Gerzon, M. A. (1973). Periphony: With-Height Sound Reproduction. *Journal of the Audio Engineering Society, JAES*, 21(1).

Gerzon, M. A. (1985). Ambisonics in Multichannel Broadcasting and Video. *Journal of the Audio Engineering Society, JAES*, 33(11).

Gilkey, R. and Anderson, T. R. (1997). *Binaural and Spatial Hearing in Real and Virtual Environments.* Lawrence Erlbaum Associates Inc.

Goertz, A. (2008). *Handbuch der Audiotechnik, Lautsprecher.* Editor: Stefan Weinzierl. Springer.

Guldenschuh, M., Sontacchi, A., and Zotter, F. (2008). HRTF Modelling in Due Consideration Variable Torso Reflections. *Proceedings of the Acoustics.*

Gumerov, N. A. and Duraiswami, R. (2004). *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions.* Elsevier.

Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N. (2010). Computation of the Head-Related Transfer Function via the Fast Multipole Accelerated Boundary Element Method and its Spherical Harmonic Representation. *Journal of the Acoustical Society of America, JASA*, 127(1).

Helwani, K., Buchner, H., and Spors, S. (2011). Calibration of Microphone Arrays with Arbitrary Geometries. *Proceedings of the German DAGA Conference, DEGA.*

Herre, J., Hilpert, J., Kuntz, A., and Plosties, J. (2014). MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding. *Journal of the Audio Engineering Society, JAES*, 62(12).

Hess, W. (2012). Head-Tracking Techniques for Virtual Acoustics Applications. *Proceedings of the AES Convention*, 133, Paper 8782.

Hom, R. C.-M., Algazi, V. R., and Duda, R. O. (2006). High-Frequency Interpolation for Motion- Tracked Binaural Sound. *Proceedings of the AES Convention*, 121, Paper 6963.

Hu, H., Zhou, L., Zhang, J., Ma, H., and Wu, Z. (2006). Head Related Transfer Function Personalization Based on Multiple Regression Analysis. *Proceedings of the International Conference on Computational Intelligence and Security, IEEE.*

ISO (2003). ISO 3745:2003(E): Acoustics - Determination of Sound Power Levels of Noise Sources Using Sound Pressure - Precision Methods for Anechoic and Hemi - Anechoic Rooms.

ITU (1994). Rec. ITU-R BS.1116-1: Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems.

ITU (2003). Rec. ITU-R BS.1534-1: Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems.

ITU (2014). Rec. ITU-R BS.1534-2: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems.

Jackson, J. D. (1962). *Classical electrodynamics*. John Wiley & Sons.

Jackson, L. B. (1996). *Digital Filters and Signal Processing*. Springer.

Jeffet, M. and Rafaely, B. (2014). Study of a Generalized Spherical Array Beamformer with Adjustable Binaural Reproduction. *Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays HSCMA*.

Jin, J.-M. (2011). *Theory and Computation of Electromagnetic Fields*. Wiley-IEEE Press.

Karamustafaoglu, A., Horbach, U., Pellegrini, R., Mackensen, P., and Theile, G. (1999). Design and Applications of a Data-Based Auralization System for Surround Sound. *Proceedings of the AES Convention*, 106, Paper 4976.

Katz, B. F. G. (2001a). Boundary Element Method Calculation of Individual Head-Related Transfer Function. I. Rigid Model Calculation. *Journal of the Acoustical Society of America, JASA*, 110(5).

Katz, B. F. G. (2001b). Boundary Element Method Calculation of Individual Head-Related Transfer Function. II. Impedance Effects and Comparisons to Real Measurements. *Journal of the Acoustical Society of America, JASA*, 110(5).

Katz, B. F. G. and Parseihian, G. (2012). Perceptually Based Head-Related Transfer Function Database Optimization. *Journal of the Acoustical Society of America, JASA*, 131(2).

Kefauver, A. P. and Patschke, D. (2007). *Fundamentals of Digital Audio*. A-R Editions.

Kennedy, R. A., Abhayapala, T., Ward, D. B., and Williamson, R. C. (1996). Nearfield Broadband Frequency Invariant Beamforming. *Proceedings of the ICASSP, IEEE*.

Kock, W. E. (1950). Binaural Localization and Masking. *Journal of the Acoustical Society of America, JASA*, 22(6).

Koretz, A. and Rafaely, B. (2009). Dolph-Chebyshev Beampattern Design for Spherical Arrays. *Transactions on Signal Processing, IEEE*, 57(6).

Koyama, S., Shimauchi, S., and Ohmuro, H. (2014). Sparse Sound Field Representation in Recording and Reproduction for Reducing Spatial Aliasing Artifacts. *Proceedings of the ICASSP, IEEE*.

Krokstad, A., Strøm, S., and Sørsdal, S. (1968). Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique. *Journal of Sound and Vibration*, 8(1).

Kuttruff, H. (2004). *Akustik: Eine Einführung*. Hirzel.

Lebedev, V. I. (1977). Spherical Quadrature Formulas Exact to Orders 25–29 . *Siberian Mathematical Journal*, 18(1).

Lentz, T. (2006). Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments. *Journal of the Audio Engineering Society, JAES*, 54.

Li, Z. and Duraiswami, R. (2007). Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming. *Transactions on Audio, Speech, and Language Processing, IEEE*, 15(2).

Li, Z. and Ruraiswami, R. (2005). Hemispherical Microphone Arrays for Sound Capture and Beamforming. *Proceedings of the Workshop on the Applications of Signal Processing to Audio and Acoustics, IEEE*.

Lindau, A. (2014a). Binaural Resynthesis of Acoustical Environments. Technology and Perceptual Evaluation. *PhD Thesis, Technical University of Berlin*.

Lindau, A. (2014b). Spatial Audio Quality Inventory (SAQI). Test Manual. *Technical University of Berlin*.

Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkmann, F., and Weinzierl, S. (2014). A Spatial Audio Quality Inventory (SAQI). *Acta Acustica united with Acustica*, 100(5).

Lindau, A., Estrella, J., and Weinzierl, S. (2010). Individualization of Dynamic Binaural Synthesis by Real Time Manipulation of ITD. *Proceedings of the AES Convention*, 128, Paper 8088.

Lindau, A. and Roos, S. (2010). Perceptual Evaluation of Discretization and Interpolation for Motion-Tracked Binaural (MTB) Recordings. *Proceedings of the VDT International Convention*.

Lindau, A. and Weinzierl, S. (2006). FABIAN - An Instrument for Software-Based Measurement of Binaural Room Impulse Responses in Multiple Degrees of Freedom. *Proceedings of the VDT International Convention*.

Mackensen, P. (2004). Auditive Localization. Head Movements, an Additional Cue in Localization. *PhD Thesis, Technical University of Berlin*.

Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M. (2013). Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions. *Proceedings of the AES Convention*, 134, Paper 8880.

McEwen, J. D., Puy, G., Thiran, J.-P., Vandergheynst, P., Van De Ville, D., and Wiaux, Y. (2013). Sparse Image Reconstruction on the Sphere: Implications of a New Sampling Theorem. *Transactions on Image Processing, IEEE*, 22(6).

Melchior, F. (2011). Investigations on Spatial Sound Design Based on Measured Room Impulse Responses. *PhD Thesis, Technical University of Delft.*

Melchior, F., Marston, D., Pike, C., Satongar, D., and Lam, Y. W. (2014). A Library of Binaural Room Impulse Responses and Sound Scenes for Evaluation of Spatial Audio Systems. *Proceedings of the German DAGA Conference, DEGA.*

Melchior, F., Thiergart, O., Galdo, G., Vries, D. d., and Brix, S. (2009). Dual Radius Spherical Cardioid Microphone Arrays For Binaural Auralization. *Proceedings of the AES Convention*, 127, Paper 7855.

Melick, J. B., Algazi, V. R., Duda, R. O., and Thompson, D. M. (2004). Customization for Personalized Rendering of Motion-Tracked Binaural Sound. *Proceedings of the AES Convention*, 117, Paper 6225.

Mellert, V. and Tohtuyeva, N. (1997). Multimicrophone Arrangement as a Substitute for Dummy-Head Recording Technique. *Journal of the Acoustical Society of America, JASA*, 102(5).

Menzel, D., Wittek, H., Theile, G., and Fastl, H. (2006). Binaurale Raumsynthese mittels Wellenfeldsynthese - Realisierung und Evaluierung. *Proceedings of the German DAGA Conference, DEGA.*

Meuleman, J. (2011). Entwurf und Aufbau eines konzentrischen Mehrwegedodekaeders. *Diploma Thesis, Cologne University of Applied Sciences.*

Meuleman, J., Bernschütz, B., and Pörschmann, C. (2011). Entwurf und Aufbau eines konzentrischen Mehrwegedodekaeders. *Proceedings of the German DAGA Conference, DEGA.*

Meyer, J. and Elko, G. (2002). A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield. *Proceedings of the ICASSP, IEEE.*

Meyer, J. and Elko, G. W. (2008). Handling Spatial Aliasing in Spherical Array Applications. *Proceedings of the Workshop on Hands-Free Speech Communication and Microphone Arrays HSCMA, IEEE.*

Middlebrooks, J. C. (1999). Virtual Localization Improved by Scaling Nonindividualized External-Ear Transfer Functions in Frequency. *Journal of the Acoustical Society of America, JASA*, 106(3).

Møller, H. (1992). Fundamentals of Binaural Technology. *Applied Acoustics*, 36(3).

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (1996). Binaural Technique: Do We Need Individual Recordings? *Journal of the Audio Engineering Society, JAES*, 44(6).

Morrison, M. A. and Parker, G. A. (1987). A Guide to Rotations in Quantum Mechanics. *Australian Journal of Physics*, 40(4).

Möser, M. (2007). *Technische Akustik*. Springer.

Muhammad, I., Jang, H. S., and Jeon, J. Y. (2014). Virtual Sound Field Immersions by Beamforming and Effective Crosstalk Cancellation Using Wavelet Transform Analysis. *Proceedings of the Forum Acusticum, EAA*.

Müller, S. (1999). Digitale Signalverarbeitung für Lautsprecher. *PhD Thesis, RWTH Aachen University*.

Nelson, P. A. and Kahana, Y. (2001). Spherical Harmonics, Singular-Value Decomposition and the Head-Related Transfer Function. *Journal of Sound and Vibration*, 239(4).

Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, 47(2).

O'Donovan, A., Zotkin, D. N., and Duraiswami, R. (2008). A Spherical Microphone Array Based System for Immersive Audio Scene Rendering. *Technical Reports of the Computer Science Department, UMIACS*.

Ohm, J.-R. and Lüke, H. D. (2004). *Signalübertragung: Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme*. Springer.

Pail, R., Plank, G., and Schuh, W. D. (2001). Spatially Restricted Data Distributions on the Sphere: The Method of Orthonormalized Functions and Applications. *Journal of Geodesy*, 75(1).

Paul, S. (2009). Binaural Recording Technology: A Historical Review and Possible Future Developments. *Acta Acustica united with Acustica*, 95(5).

Pelzer, S., Pollow, M., and Vorländer, M. (2012). Continuous and Exchangeable Directivity Patterns in Room Acoustic Simulation. *Proceedings of the German DAGA Conference, DEGA*.

Pendleton, J. D. (2003). Euler Angle Geometry, Helicity Basis Vectors, and the Wigner D-Function Addition Theorem. *American Journal of Physics*, 71(12).

Perrett, S. and Noble, W. (1997). The Effect of Head Rotations on Vertical Plane Sound Localization. *Journal of the Acoustical Society of America, JASA*, 102(4).

Peters, N. and Schmeder, A. W. (2011). Beamforming Using a Spherical Microphone Array Based on Legacy Microphone Characteristics. In *Proceedings of the ICSA International Conference on Spatial Audio*.

Plenge, G. (1974). On the Differences Between Localization and Lateralization. *Journal of the Acoustical Society of America, JASA*, 56.

Pollack, I. and Rose, M. (1967). Effect of Head Movement on the Localization of Sounds in the Equatorial Plane. *Perception and Psychophysics*, 2(12).

Pollow, M. (2010). Applying Extrapolation and Interpolation Methods to Measured and Simulated HRTF Data Using Spherical Harmonic Decomposition. *Proceedings of the International Symposium on Ambisonics and Spherical Acoustics, EAA*.

Pollow, M., Klein, J., Dietrich, P., Behler, G. K., and Vorlaender, M. (2012). Optimized Spherical Sound Source for Room Reflection Analysis. *Proceedings of the International Workshop on Acoustic Signal Enhancement IWAENC, VDE*.

Rafaely, B. (2004). Plane-Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution. *Journal of the Acoustical Society of America, JASA*, 116(4).

Rafaely, B. (2005). Analysis and Design of Spherical Microphone Arrays. *Transactions on Speech and Audio Processing, IEEE*, 13(1).

Rafaely, B. (2015). *Fundamentals of Spherical Array Processing*. Springer.

Rafaely, B., Balmages, I., and Eger, L. (2007a). High-Resolution Plane-Wave Decomposition in an Auditorium Using a Dual-Radius Scanning Spherical Microphone Array. *Journal of the Acoustical Society of America, JASA*, 122(5).

Rafaely, B. and Kleider, M. (2008). Spherical Microphone Array Beam Steering Using Wigner-D Weighting. *Signal Processing Letters, IEEE*.

Rafaely, B., Weiss, B., and Bachmat, E. (2007b). Spatial Aliasing in Spherical Microphone Arrays. *Transactions on Signal Processing, IEEE*, 55(3).

Rasumow, E., Blau, M., Doclo, S., Hansen, M., van de Par, S., Püschel, D., and Mellert, V. (2013). Least Squares Versus Non-Linear Cost Functions for a Virtual Artificial Head. *Journal of the Acoustical Society of America, JASA*, 133(5).

Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Mellert, V., and Püschel, D. (2014a). The Impact of the White Noise Gain (WNG) of a Virtual Artificial Head on the Appraisal of Binaural Sound Reproduction. *Proceedings of the Joint Symposium on Auralization and Ambisonics, EAA*.

Rasumow, E., Blau, M., Hansen, M., Doclo, S., van der Par, S., Mellert, V., and Püschel, D. (2011). Robustness of Virtual Artifcial Head Topologies with Respect to Microphone Positioning. In *Proceedings of the Forum Acusticum, EAA*.

Rasumow, E., Blau, M., Hansen, M., van de Par, S., Doclo, S., Mellert, V., and Püschel, D. (2014b). Smoothing Individual Head-Related Transfer Functions in the Frequency and Spatial Domains. *Journal of the Acoustical Society of America, JASA*, 135(4).

Rauhut, H. and Ward, R. (2011). Sparse Recovery for Spherical Harmonic Expansions. *math.NA arXiv*, 1102.4097.

Reilly, A. and McGrath, D. (1995). Real-Time Auralization with Head Tracking. *Proceedings of the Australian Regional AES Convention*, 5, Paper 4024.

Rettberg, T., Helwani, K., Spors, S., and Buchner, H. (2012). Practical Aspects of the Calibration of Spherical Microphone Arrays. *Proceedings of the German DAGA Conference, DEGA*.

Rettberg, T. and Spors, S. (2013). On the Impact of Noise Introduced by Spherical Beamforming Techniques on Data-Based Binaural Synthesis. *Proceedings of the German DAGA Conference, DEGA*.

Rettberg, T. and Spors, S. (2014). Time-Domain Behaviour of Spherical Microphone Arrays at High Orders. *Proceedings of the German DAGA Conference, DEGA*.

Rühl, M. (2012). Entwicklung eines Verfahrens zur Detektion von Reflexionen in messtechnisch erfassten Schallfeldern. *Bachelor Thesis, Cologne University of Applied Sciences*.

Sakamoto, S., Kodama, J., Hongo, S., Okamoto, T., Iwaya, Y., and Suzuki, Y. (2010). A 3D Sound-Space Recording System Using Spherical Microphone Array with 252Ch Microphones. *Proceedings of Meetings on Acoustics ICA, ASA*.

Salvador Castaneda, C. D., Sakamoto, S., Trevino Lopez, J. A., Li, J., Yan, Y., and Suzuki, Y. (2013). Accuracy of Head-Related Transfer Functions Synthesized with Spherical Microphone Arrays. *Proceedings of Meetings on Acoustics ICA, ASA*.

Satongar, D., Lam, Y. W., and Pike, C. (2014). Measurement and Analysis of a Spatially Sampled Binaural Room Impulse Response Dataset. *Proceedings of the International Congress on Sound and Vibration ICSV*.

Schärer, Z. and Lindau, A. (2012). Evaluation of Equalization Methods for Binaural Signals. *Proceedings of the AES Convention*, 126, Paper 7721.

Schlesinger, A., Albrecht, B., Galdo, G. D., and Husung, S. (2007). Holographic Sound Field Analysis with a Scalable Spherical Microphone Array. *Proceedings of the AES Convention*, 122, Paper 7145.

Schultz, F. and Spors, S. (2013). Data-Based Binaural Synthesis Including Rotational and Translatory Head-Movements. *Proceedings of the AES Conference on Sound Field Control - Engineering and Perception*, 53, Paper P7.

Seeber, B. U. and Fastl, H. (2003). Subjective Selection of Non-Individual Head-Related Transfer Functions. *Proceedings of the International Conference on Auditory Display ICAD*.

Shabtai, N. R. and Rafaely, B. (2012). Spherical Array Beamforming for Binaural Sound Reproduction. *Convention of Electrical and Electronics Engineers, IEEE*.

Shabtai, N. R. and Rafaely, B. (2013a). Binaural Sound Reproduction Beamforming Using Spherical Microphone Arrays. *Proceedings of the ICASSP, IEEE*.

Shabtai, N. R. and Rafaely, B. (2013b). Spherical Array Processing with Binaural Sound Reproduction for Improved Speech Intelligibility. *Proceedings of Meetings on Acoustics ICA, ASA*.

Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IEEE*, 86(2).

Sheaffer, J. and Rafaely, B. (2014). Equalization Strategies For Binaural Room Impulse Response Rendering Using Spherical Arrays. *Proceedings of the IEEEI Convention, IEEE*.

Sheaffer, J., van Walstijn, M., Rafaely, B., and Kowalczyk, K. (2014a). A Spherical Array Approach for Simulation of Binaural Impulse Responses using the Finite Difference Time Domain Method. *Proceedings of the Forum Acusticum, EAA*.

Sheaffer, J., Villeval, S., and Rafaely, B. (2014b). Rendering Binaural Room Impulse Responses from Spherical Microphone Array Recordings Using Timbre Correction. *Proceedings of the Joint Symposium on Auralization and Ambisonics, EAA*.

Snyder, J. P. (1997). *Flattening the Earth: Two Thousand Years of Map Projections*. University of Chicago Press.

Song, W., Ellermeier, W., and Hald, J. (2011). Psychoacoustic Evaluation of Multichannel Reproduced Sounds Using Binaural Synthesis and Spherical Beamforming. *Journal of the Acoustical Society of America, JASA*, 130(4).

Sporer, T., Liebetrau, J., and Schneider, S. (2009). Statistics of MUSHRA Revisited. *Proceedings of the AES Convention*, 127, Paper 7825.

Spors, S. (2006). Active Listening Room Compensation for Spatial Sound Reproduction Systems. *PhD Thesis, Friedrich-Alexander-University Erlangen-Nürnberg*.

Spors, S., Rabenstein, R., and Ahrens, J. (2008). The Theory of Wave Field Synthesis Revisited . *Proceedings of the AES Convention*, 124, Paper 7358.

Spors, S. and Wierstorf, H. (2012). Evaluation of Perceptual Properties of Phase-mode Beamforming in the Context of Data-based Binaural Synthesis. *Proceedings of the International Symposium on Communications Control and Signal Processing ISCCSP, IEEE*.

Spors, S., Wierstorf, H., and Ahrens, J. (2012a). Interpolation and Range Extrapolation of Head-Related Transfer Functions using Virtual Local Wave Field Synthesis. *Proceedings of the AES Convention*, 130, Paper 8392.

Spors, S., Wierstorf, H., and Geier, M. (2012b). Comparison of Modal Versus Delay-And-Sum Beamforming in The Context of Data-Based Binaural Synthesis. *Proceedings of the AES Convention*, 132, Paper 8669.

Stade, P. (2013). Zur Untersuchung der Akustik von Studios im WDR-Funkhaus unter Anwendung von Wellenfeldanalyseverfahren. *Master Thesis, Cologne University of Applied Sciences*.

Stade, P., Bernschütz, B., and Rühl, M. (2012). A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios. *Proceedings of the VDT International Convention*.

Stroud, A. H. and Secrest, D. (1966). *Gaussian Quadrature Formulaes*. Prentice-Hall.

Sun, H., Teutsch, H., Mabande, E., and Kellermann, W. (2011). Robust Localization of Multiple Sources in Reverberant Environments Using EB-Esprit with Spherical Microphone Arrays. *Proceedings of the ICASSP, IEEE*.

Teutsch, H. (2007). *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Springer.

Theile, G. (1986a). Das Kugelflächenmikrofon. *Proceedings of the VDT International Convention*.

Theile, G. (1986b). On the Standardization of the Frequency Response of High-Quality Studio Headphones. *Journal of the Audio Engineering Society*, 34(12).

Thompson, B. (2006). *Foundations of Behavioral Statistics: An Insight-Based Approach*. Guilford Press.

Tohtuyeva, N. and Mellert, V. (1999). Approximation of Dummy-Head Recording Technique by a Multimicrophone Arrangement. *Journal of the Acoustical Society of America, JASA*, 105(2).

Toshima, I., Uematsu, H., and Hirahara, T. (2003). A Steerable Dummy Head that Tracks Three-Dimensional Head Movement: TeleHead. *Acoustical Science and Technology*, 24(5).

Varshalovich, D. A., Moskalev, A. N., and Khersonskii, V. K. (1988). *Quantum Theory of Angular Momentum*. World Scientific.

Vázquez-Giner, A. (2013). Scale: A Software Tool for Listening Experiments. *Proceedings of the German DAGA Conference, DEGA*.

Vázquez-Giner, A. (2015). Scale - Conducting Psychacoustic Experiments with Dynamic Binaural Synthesis. *Proceedings of the German DAGA Conference, DEGA*.

Vorländer, M. (2008). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer.

Wallach, H. (1940). The Role of Head Movements and Vestibular and Visual Cues in Sound Localization. *Journal of Experimental Psychology*, 27(4).

Weinzierl, S., Lindau, A., Brandenburg, K., de Vries, D., Maempel, H. J., van der Par, S., Rafaely, B., Spors, S., and Vorländer, M. (2012). The SEACEN Project. *Proceedings of the German DAGA Conference, DEGA*.

Weinzierl, S. and Maempel, H.-J. (2012). Sind Hörversuche subjektiv? Zur Objektivität akustischer Maße. *Proceedings of the German DAGA Conference, DEGA*.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization Using Nonindividualized Head-Related Transfer Functions. *Journal of the Acoustical Society of America, JASA*, 94.

Wenzel, E. M., Fisher, S. S., Stone, P. K., and Foster, S. H. (1990). A System for Three-Dimensional Acoustic Visualization in a Virtual Environment Workstation. *Proceedings of the Conference on Visualization, IEEE*.

Wightman, F. L. and Kistler, D. J. (1999). Resolution of Front–Back Ambiguity in Spatial Hearing by Listener And Source Movement. *Journal of the Acoustical Society of America, JASA*, 105(5).

Wigner, E. P. (1931). *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*. Vieweg.

Williams, E. (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press.

Winter, F., Schulz, F., and Spors, S. (2014). Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements. *Proceedings of the Forum Acusticum, EAA*.

Xie, B. (2009). On The Low Frequency Characteristics Of Head-Related Transfer Function. *Chinese Journal of Acoustics*, 28(2).

Young, P. T. (1931). The Role of Head Movements in Auditory Localization. *Journal of Experimental Psychology*, 14(2).

Zhang, W., Abhayapala, T. D., Kennedy, R. A., and Duraiswami, R. (2009). Modal Expansion of HRTFs: Continuous Representation in Frequency-Range-Angle. In *Proceedings of the ICASSP, IEEE*.

Zhang, W., Zhang, M., Kennedy, R. A., and Abhayapala, T. D. (2012). On High-Resolution Head-Related Transfer Function Measurements: An Efficient Sampling Scheme. *Transactions on Audio, Speech, and Language Processing, IEEE*, 20(2).

Zielinski, S., Hardisty, P., Hummersone, C., and Rumsey, F. (2007). Potential Biases in MUSHRA Listening Tests. *Proceedings of the AES Convention*, 123, Paper 7179.

Zielinski, S., Rumsey, F., and Bech, S. (2008). On Some Biases Encountered in Modern Audio Quality Listening Tests - A Review . *Journal of the Audio Engineering Society, JAES*, 56(6).

Zölzer, U. (2008). *Digital Audio Signal Processing*. John Wiley & Sons.

Zotkin, D., Hwang, J., Duraiswaini, R., and Davis, L. S. (2003). HRTF Personalization Using Anthropometric Measurements. *Proceedings of the Workshop on the Applications of Signal Processing to Audio and Acoustics, IEEE*.

Zotkin, D. N., Duraiswami, R., and Gumerov, N. A. (2009). Regularized HRTF Fitting Using Spherical Harmonics. *Proceedings of the Workshop on the Applications of Signal Processing to Audio and Acoustics, IEEE*.

Zotter, F. (2009a). Analysis and Synthesis of Sound-Radiation with Spherical Arrays. *PhD Thesis, University of Music and Performing Arts, Graz*.

Zotter, F. (2009b). Sampling Strategies for Acoustic Holography/Holophony on the Sphere. *Proceedings of the NAG-DAGA Conference, AIA-DEGA*.