CHALMERS
UNIVERSITY OF TECHNOLOGY

Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data

Tim Lübeck
Department of Civil Engineering and Architecture
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis 2019

Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data

Tim Lübeck

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
Gothenburg, Sweden 2019

Technology
Arts Sciences
TH Köln

Faculty of Information-, Media- and Electrical Engineering
Institute of Communications Engineering
TH Köln - University of Applied Sciences

Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data
Tim Lübeck

Supervisor:
Jens Ahrens, Chalmers University of Technology, Department of Architecture and Civil Engineering

Examiner:
Jens Ahrens, Chalmers University of Technology, Department of Architecture and Civil Engineering
Christoph Pörschmann, TH Köln, Institute of Communications Engineering

Master's Thesis ACEX30-19-101
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Python simulation of a plane wave impact on a spherical microphone array.

Typeset in LaTeX
Gothenburg, Sweden 2019

Perceptual Evaluation of Mitigation Approaches of Errors due to Spatial Undersampling in Binaural Renderings of Spherical Microphone Array Data

Tim Lübeck
Department of Architecture and Civil Engineering
Chalmers University of Technology

Home University:
TH Köln - University of Applied Sciences,
Institute of Communications Engineering

Abstract

High fidelity virtual acoustic environments (VAEs) have become of increasing interest in the context of immersive virtual and augmented reality applications, and have developed to a reputable field of research. To present a real-life sound scene as VAE to a listener, the spatial sound field has to be captured. Traditionally, spatial sound fields are sequentially measured by means of a dummy head. Using spherical microphone array recordings is a promising alternative to time-consuming dummy head measurements. This enables the possibility of dynamic real-time applications that involve recordings and reproductions of spatial sound scenes as for example live concerts or telephone conferences. However, the quality of the rendered VAEs is significantly limited by the density of the microphone distribution provided by the microphone array. A limited number of microphones leads to spatial undersampling of the surrounding sound field and creates audible artifacts in the VAE. This work investigates the errors due to spatial undersampling and state-of-the-art approaches to mitigate them. In a concluding listening experiment, the improvements that can be achieved with these approaches were perceptually evaluated. It was found that there are a few algorithms that significantly improve the binaural auralization of spherical microphone array data.

Keywords: Spatial Aliasing, SH Order Truncation, Spherical Harmonics, Plane Wave Decomposition, Spherical Microphone Arrays

Acknowledgements

First of all, I would like to thank my supervisor Jens Ahrens whose door was always open to me and my questions. Through him, I have gained a lot of expertise for my future career.
Many thanks go to my second supervisor Christoph Pörschmann. He already accompanies me throughout my entire studies and played a decisive role in my current graduation. He also opens up important chances for my future personal and career development.
I also would like to thank the entire Division of Applied Acoustics of Chalmers University for giving me the opportunity to work on this project. It was always a pleasant atmosphere, that made my time at Chalmers a great experience and a perfect completion of my university studies. At this stage, I want to say thank you to all the people at the Division who participated in my final listening experiment. Further, I would like to thank my office colleague Hannes Helmholz for the enjoyable cooperation. Hopefully, there will be more common projects in the future.
Last but not least, I would like to thank Johannes Arend who taught me all my basic knowledge in audio processing and scientific writing and still has the patience to guide me with the statistical evaluations.
As this work was done during an Erasmus semester abroad at the Chalmers University, many thanks go to the Erasmus program for their financial support.

Tim Lübeck, Gothenburg, August 2019

Contents

1
Introduction

1.1 Motivation

In the last decades the rising number of virtual and augmented reality applications creates the demand for high fidelity virtual acoustic environments (VAEs). Furthermore, presenting immersive spatial sound fields to a listener has become a reputable field of research. One common way of presenting a VAE is binaural synthesis. The headphone-based auralization stimulates the binaural spatial hearing of the human auditory system, and ideally reproduces the same sound pressure at the ear drums that would be perceived in a real-life sound scene. Binaural renderings are based on appropriate head-related impulse responses (HRIRs) or binaural room impulse responses (BRIRs) that describe the transformation of sound reaching the human ear drums. This information is usually obtained by a dummy head with two microphones in the ear canals which capture the surrounding sound field. For presenting a full-spherical VAE, in particular when involving the head-orientation of the listener, time-consuming dummy head measurements with adequately high resolution are unavoidable.
When thinking of dynamic auralizations of spatial sound fields, for example broadcastings of live concerts, or spatial acoustic telephone conferences, dummy head measurements are unfeasible. An alternative to dummy head measurements is the recording of spatial sound fields by means of spherical microphone arrays. Microphone arrays allow to capture the entire surrounding sound field at once, instead of sequentially measuring it with a dummy head. The reproduction of array recordings is not restricted to binaural auralizations, but could also involve wave field syntheses, Ambisonic reproductions or any other object based VAE. Furthermore, spherical microphone arrays can be used for a wide range of further applications and research areas, for example sound field analyses as beamforming or acoustic holography.
This work focuses on binaural reproductions of spherical microphone array capturings. Bernschütz (2016) elaborated a mathematical closed description for recording a spatial sound field by means of spherical microphone arrays, and reproducing it via dynamic binaural synthesis. Theoretically, this reproduction works without any loss of quality compared to dummy head auralizations. However, real-life implementations have technical constrains, mainly caused by a limited number of microphones that lead to spatial undersampling of the surrounding sound field. Two technical constraints, namely the spatial aliasing and order truncation effect, lead to audible artifacts in the binaural rendering. They are the spatial analogous to the artifacts occurring at the time-frequency sampling.

This work investigates the spatial aliasing and order truncation effect and their consequences on the perceived quality of spatial audio reproductions. Moreover, state-of-the-art algorithms to reduce the influence of these artifacts are studied, implemented and compared in a conclusive listening experiment.

1.2 Related Works

Benjamin Bernschütz developed a comprehensve signal processing chain including the microphone array capturing of spatial sound fields and the reproduction of binaural signals as part of his doctoral dissertation (Bernschütz, 2016). Most of the theoretical derivations in this work are based on this dissertation. The Sound Field Analysis MATLAB toolbox (SOFiA) provides the implementation of the signal processing described in his thesis, see Bernschütz et al. (2011). The Python port[1] is an equivalent implementation realized by Christoph Hohnerlein, see Hohnerlein and Ahrens (2017). All implementations, visualizations, and the final generation of the listening experiment stimuli is done with tools from this Python Sound Field Analysis toolbox (SFA).
The auralization of the binaural signals is done with the SoundScape Renderer (SSR), developed by Geier et al. (2008). It allows to render arbitrary HRIR or BRIR data sets defined on a pure horizontal sampling grid with 1° resolution.
To achieve the main goal of spherical microphone array renderings, namely the real-time auralization, Helmholz et al. (2019) developed a Python application which allows to dynamically capture and reproduce spatial sound fields. Although this work just focuses on impulse response based renderings, it is intended to use the findings to improve this real-time auralization as much as possible.

[1] `https://github.com/AppliedAcousticsChalmers/sound_field_analysis-py`

2

Theory

2.1 Mathematical Derivation of Capturing and Reproducing Spatial Sound Fields

This chapter briefly summarizes the theory of capturing spatial sound fields with spherical microphone arrays and reproducing these captures as a virtual acoustic environment. Therefore, in the first section, the mathematical fundamentals of sound field capturing are explained. Section 2.2 describes the limitations that arise when applying the mathematical derivations to real-life spherical microphone arrays. Finally, the consequences of these technical constraints are discussed, and a number of state-of-the-art reduction approaches are introduced.

2.1.1 The Acoustic Wave Equation and its Solutions

In linear acoustics every sound pressure $p(x,t)$ in the free space x at the time t is satisfying the homogeneous acoustic wave equation, with c denoting the speed of sound in the air and ∇^2 the Laplacian in Cartesian coordinates.

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

This holds for every homogeneous fluid with no viscosity, see e.g. Kinsler et al. (1999) or Williams (1999, Eq. 2.1, p. 15). For a sound field consisting of a single-frequency, the pressure $p(x,t)$ can be expressed as

$$p(x,t) = p(x)\mathrm{e}^{\mathrm{j}\omega t}, \tag{2.2}$$

with the angular frequency $\omega = 2\pi f$ and the temporal frequency f. Using this description, the time-frequency transform of the wave equation yields the Helmholtz equation

$$\nabla^2 p(x,\omega) + \left(\frac{\omega}{c}\right)^2 p(x,\omega) = 0 \tag{2.3}$$

which delivers stationary solutions in the frequency domain. ω/c is mostly denoted as the wave number k.
As in this work spatial sound fields are measured by means of spherical microphone arrays, it is reasonable to use spherical coordinates to describe the point $r = (r, \phi, \theta)$ in the three dimensional space. Whereby ϕ denotes the azimuth angle ranging from 0 to 2π and θ the colatitude ranging from 0 to π. Expressing the

Helmholtz equation in spherical coordinates and using the assumption of a separable solution ($p(r, \phi, \theta) = R(r) \cdot \Phi(\phi) \cdot \Theta(\theta)$), the Helmholtz equation can be decomposed into three separate equations

$$\frac{d}{dr}\left(r^2 \frac{dR}{dr}\right) + \left[\left(\frac{\omega}{c}\right)^2 r^2 - n(n+1)\right] R = 0 \tag{2.4}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta \frac{d\Theta}{d\theta}\right) + \left[n(n+1) - \frac{m^2}{\sin^2\theta}\right]\Theta = 0 \tag{2.5}$$

$$\frac{d^2\Phi}{d\phi^2} + m^2\Phi = 0. \tag{2.6}$$

A detailed derivation can be found in Rafaely (2015, pp. 31-34). These equations depend either on r, ϕ, or θ only and their solutions are well known. Equation 2.4 describes the radial behavior and can be transformed to the spherical Bessel equation, solved by the spherical Bessel functions. Equation 2.5 and Equation 2.6 describe the angular components with respect to ϕ and θ and can be solved by the spherical harmonics.

2.1.1.1 Spherical Bessel Functions

With respect to the present problem, the radial behavior of the wave equation can be described by n-th order solutions. There are Bessel functions of the first kind j_n used for interior problems, whereby all sources are surrounding the array surface, or Bessel functions of the second kind denoted as y_n. Furthermore, there are Hankel functions of the first kind $h_n^{(1)} = j_n(z) + iy(z)$, or second kind $h_n^{(2)} = j_n(z) - iy_n(z)$ used for exterior problems, where the sensors completely enclose the sources to be captured. Hankel functions of the second kind are also denoted as third kind Bessel functions.

2.1.1.2 Solutions of the Angular Components - Spherical Harmonics

Equation 2.5 is denoted as Legendre equation which can be solved by associated n-th order and m-th mode Legendre functions $P_n^m(\cos\theta)$ of the first kind. Equation 2.6 is an ordinary second order differential equation which dependents on a specific integer constant m, resulting from the periodicity of Φ, as a function over the unit circle.
Both angular components of the Helmholtz equation $\Theta(\theta)$ and $\Phi(\phi)$ are usually described by the spherical harmonics (SH)

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) \mathrm{e}^{\mathrm{j}m\phi}. \tag{2.7}$$

There exist several ways of defining the spherical harmonics, see e.g. Williams (1999, Eq. 6.20, p. 186), or Jackson (1962, Eq. 3.53, p. 108) which, however, has no effect

on the fundamental theory. One important property of the spherical harmonics is that they set up an orthonormal basis of the sphere

$$\begin{aligned} < Y_n^m(\theta,\phi), Y_{n'}^{m'} > &= \int_0^{2\pi} \int_0^{\pi} Y_n^m(\theta,\phi) Y_{n'}^{m'}(\theta',\phi')^* \sin\theta d\theta d\phi \\ &= \delta_{nn'}(\phi-\phi')\delta_{mm'}(\cos\theta - \cos\theta'), \end{aligned} \tag{2.8}$$

where $(\cdot)^*$ denotes the complex conjugate. $\delta_{nn'}(\phi-\phi')\delta_{mm'}(\cos\theta-\cos\theta')$ is denoted as Dirac Delta or Kronecker Delta function.

2.1.2 The Spatial Fourier Transform

Spherical harmonics are a closed set of solutions for the angular component of the Laplacian and thus the Helmholtz equation. Furthermore, they form a complete set of orthogonal basis functions of the sphere. Therefore, arbitrary functions fulfilling the wave equation can be expanded over a sphere by means of the spherical harmonics. The time-frequency Fourier transform expands functions as a linear combination of the orthogonal sine and cosine functions over a circle. Similarly, the spatial or spherical Fourier transform (SFT) expands spatial functions as a linear combination of the orthogonal SHs over a sphere. The spatial Fourier transform is often denoted as spherical harmonic expansion and the domain described by the spatial Fourier expansion coefficients (SH coefficients) as spherical wave spectrum or SH domain. For an arbitrary function $G(\phi,\theta,r,\omega)$ the SH coefficients $\mathring{G}_{nm}(r,\omega)$ can be calculated with

$$\mathring{G}_{nm}(r,\omega) = \int_0^{2\pi} \int_0^{\pi} G(r,\phi,\theta,\omega) Y_n^m(\theta,\phi)^* \sin\theta d\theta d\phi. \tag{2.9}$$

During the entire work, $\mathring{(\cdot)}$ is denoting the SH coefficients in spherical coordinates. The inverse spatial Fourier transform (ISFT) is given by

$$G(r,\phi,\theta,\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{G}_{nm}(r,\omega) Y_n^m(\theta,\phi). \tag{2.10}$$

2.1.3 The Plane Wave Composition

Theoretically, a plane wave arises from an infinitely distant point source, such that a straight plane wavefront can be assumed at the point of interest. According to Rafaely (2003), every sound field can be described as a superposition of multiple plane waves. For the further processing, it becomes useful to describe sound fields as such a plane wave continuum to find the so-called plane wave (PW) components $D(\phi_d,\theta_d,\omega)$. These PW components or PW coefficients describe which portion of a unit-amplitude plane wave impinging on a sphere at (ϕ_d,θ_d) would generate the same sound pressure as the surrounding sound field at that point.
According to Williams (1999, p. 259), the SH expansion with respect to the position (r_0,ϕ,θ) of an unit-amplitude plane wave impinging on an open sphere at (θ_d,ϕ_d) can be expressed as

$$p(r_0,\phi,\theta,\phi_d,\theta_d,\omega) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathrm{i}^n j_n\left(\frac{\omega}{c} r\right) Y_n^m(\theta,\phi) Y_n^m(\theta_d,\phi_d)^*, \tag{2.11}$$

with the imaginary unit i and the spherical Bessel function j_n. The sound field $s(r_0, \phi, \theta, \omega)$ can now be expressed as the summation of an infinite number of weighted plane waves impinging on a sensor surface S_0 with radius r_0 from all possible directions (ϕ_d, θ_d)

$$s(r_0, \phi, \theta, \omega) = \int_0^{2\pi} \int_0^{\pi} D(\phi_d, \theta_d, \omega,) p(r_0, \phi_d, \theta_d, \phi, \theta, \omega) \sin\theta d\theta d\phi, \tag{2.12}$$

where $D(\phi_d, \theta_d, \omega)$ are the desired frequency dependent weighting coefficients of the unity plane waves $p(r_0, \phi, \theta, \phi_d, \theta_d, \omega)$. Substituting Equation 2.11 in Equation 2.12 and rearanging in favor of expressing $D(\phi_d, \theta_d, \omega)$ yields

$$D(\phi_d, \theta_d, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{4\pi \mathrm{i}^n j_n\left(\frac{\omega}{c} r_0\right)} \mathring{S}_{nm}(r_0, \omega) Y_n^m(\theta_d, \phi_d), \tag{2.13}$$

where $\mathring{S}_{nm}(r_0, \omega)$ denotes the SH coefficients of the sound field $s(r_0, \omega, \phi, \theta)$. This equation is known as the conventional plane wave decomposition (PWD), see e.g. Bernschütz (2016, p. 63) or Rafaely (2003, pp. 42-45).
Considering Equation 2.11 and keeping the property of orthogonality (Eq. 2.8) in mind, the PWD can be interpreted as a varying spatial Dirac impulse in (ϕ_d, θ_d) direction, sampling the spatial sound field.
Equation 2.13 refers to a sound field at an open sphere. The term

$$d_n = \frac{1}{4\pi \mathrm{i}^n j_n\left(\frac{\omega}{c} r_0\right)} \tag{2.14}$$

is denoted as radial filter d_n and contains the spherical Bessel functions j_n, discussed in Section 2.1.1.1. These radial filters basically remove the dependency of the expansion surface and thus depend on the array setup. Expressing the PWD more generally leads to

$$D(\phi_d, \theta_d, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n \mathring{S}_{nm}(r_0, \omega) Y_n^m(\phi_d, \theta_d). \tag{2.15}$$

When calculating the PW components at a rigid sphere, the radial filters d_n have to compensate for scattering effects of the array body and thus will be extended to

$$d_n = \frac{1}{4\pi \mathrm{i}^n \left(j_n\left(\frac{\omega}{c} r_0\right) - \frac{j_n'(\frac{\omega}{c} r_0)}{h_n'^{(2)}(\frac{\omega}{c} r_0)} h_n^{(2)}(\frac{\omega}{c} r_0) \right)}. \tag{2.16}$$

Radial filters which are usually realized as FIR filters have a significant influence on the auralization quality, and have been discussed extensively in past investigations, e.g. Bernschütz (2016, pp. 90-118), Rafaely (2015, pp. 34-38) or Lösler and Zotter (2015). However, this work will not cover the radial filters in more detail [1].
Figure 2.1 shows magnitudes of PW coefficients of a simulated broadband plane wave impinging on an ideal array at $\phi = 0°$, $\theta = 90°$, with respect to the frequency and azimuth direction.

[1] Diagrams of the radial filters used for the auralization in this work are depicted in the appendix, see Figure A.1

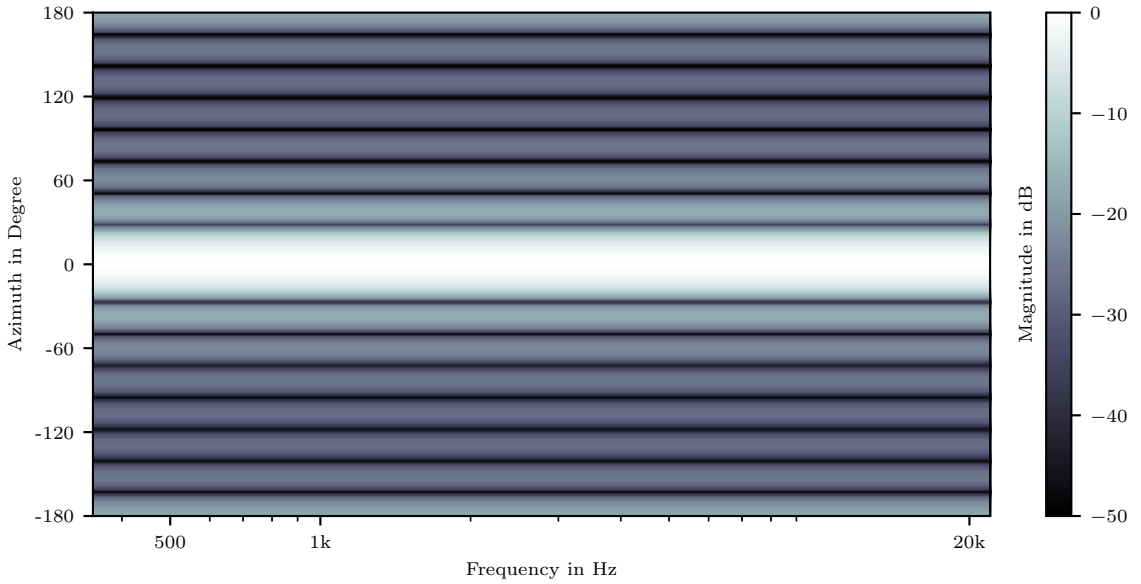

Figure 2.1: A simulated broadband plane wave impinging on an ideal array with a radius of 10 cm at $\phi = 0°$, $\theta = 90°$ direction. The figure shows the magnitudes of the plane wave coefficients for different azimuth angles at the array surface with respect to the frequency. The plane wave coefficients were rendered up to an SH order of 7.

2.1.4 Binaural Reproduction

To reproduce the spatial sound field recording as binaural VAE, the PW components have to be merged with corresponding head-related transfer functions (HRTF) in an appropriate way. An HRTF can be understood as the transformation of a single plane wave from a sound source to the human ears. Thus, the binaural signals $Y^{l,r}(\omega)$ that a sound field would generate at the eardrums can be calculated by weighting every PW coefficient D from direction (ϕ_d, θ_d) with corresponding HRTFs from the same direction and integrate them over the entire array surface

$$Y^{l,r}(\omega) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} H^{l,r}(\phi, \theta, \omega) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n \mathring{S}_{nm}(r_0, \omega) Y_n^m(\phi_d, \theta_d) \, \sin\theta d\theta d\phi. \tag{2.17}$$

$H^{l,r}(\phi, \theta, \omega)$ is denoting the HRTFs. A slightly different method for calculating the binaural signals can be found, for example, in Ben-Hur et al. (2018). The HRTF set $H^{l,r}(\phi, \theta, \omega)$ is transformed into the wave spectrum domain as well. By employing the property of orthogonality of the spherical harmonics (Eq. 2.8) Equation 2.17 can be rearranged to

$$Y^{l,r}(\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n \mathring{S}_{nm}(\omega, r_0) \mathring{H}_{nm}^{l,r}(\omega), \tag{2.18}$$

where $\mathring{H}_{nm}^{l,r}$ denotes the HRTF SH coefficients. This way of binaural reproduction highly depends on the convention of the SH basis functions which was presented in detail by Andersson (2017, p. 7).

2.2 Spherical Array Processing

2.2.1 Discretization - Sampling Scheme

So far, a continuous knowledge of the sound pressure distribution on the array surface was expected for the mathematical derivations. However, real-world microphone arrays do not provide a closed sensor surface S_0, and the capturing of a continuous sound pressure is impossible. Instead, there exist a few spherical microphone arrays with a certain number of microphones. For example some first order arrays as the Core Sound Tetra Mic, or Sennheiser AMBEO, described in Gonzalez et al. (2018), or some higher order microphone arrays as the 19 channel Zylia[2], the 32 channel EIGENMIKE (Meyer and Elko, 2002), or the planned 7th order HØSMA 7N (Dziwis et al., 2019). The positions of the microphones on the array surface, denoted as sampling points or sampling nodes, are defined by the sampling scheme. In the following, M_{sg} (sg: sampling grid) describes the number of microphones which are placed at the positions (ϕ_{sp}, θ_{sp}) (sp denotes the index of the sampling point). To every microphone, one weight w_{sp} is assigned which indicates the portion of the array surface covered by the corresponding microphone. It can be shown that the arrangement of the microphone positions has a significant influence on the following derivations, although the amount of microphones remains the same. Different sampling schemes are discussed in previous studies, for example Zotter (2009a) or Zotter (2009b). Every sampling scheme has an individual order N_{sg} dependent on the scheme efficiency η_{sg}. It can be calculated with $M_{sg} \approx \eta_{sg}(N_{sg}+1)^2$. Thus, to expand up to an SH order of N_{sg} at least M_{sg} microphones have to be provided by the microphone array.
For simplification, the following derivations just consider arrangements with an efficiency of one and microphone positions on a single radius r_0. Thus, the dependency on r_0 vanishes. Furthermore, the direction (ϕ, θ) will be denoted as Ω.

2.2.2 The Discrete Spatial Fourier Transform

Instead of a continuous SFT given by Equation 2.9, sampling at discrete microphone positions yields the discrete SFT (DSFT)

$$\mathring{G}_{nm}(\omega) = \sum_{sp=1}^{M_{sg}} w_{sp}\, G(\Omega_{sp}, \omega) Y_n^m(\Omega_{sp})^*. \tag{2.19}$$

Consequently, the limited number of sound field SH coefficients limits the modal order of the PWD to N_{sg}

$$D(\Omega_d, \omega) = \sum_{n=0}^{N_{sg}} \sum_{m=-n}^{n} d_n\, \mathring{S}_{nm}(\omega) Y_n^m(\Omega_d). \tag{2.20}$$

The ideal directional Dirac impulse becomes broader which leads to a reduced spatial resolution. Figure 2.2 shows the influence of the limitation of the PWD order. For

[2] https://www.zylia.co/

small orders N, the number of side-lobes decreases, but their magnitudes become larger. For higher orders, the number of side-lobes increases, but their magnitude levels decrease. For ideal sampling, and thus an infinite PWD order, all side-lobes vanish.

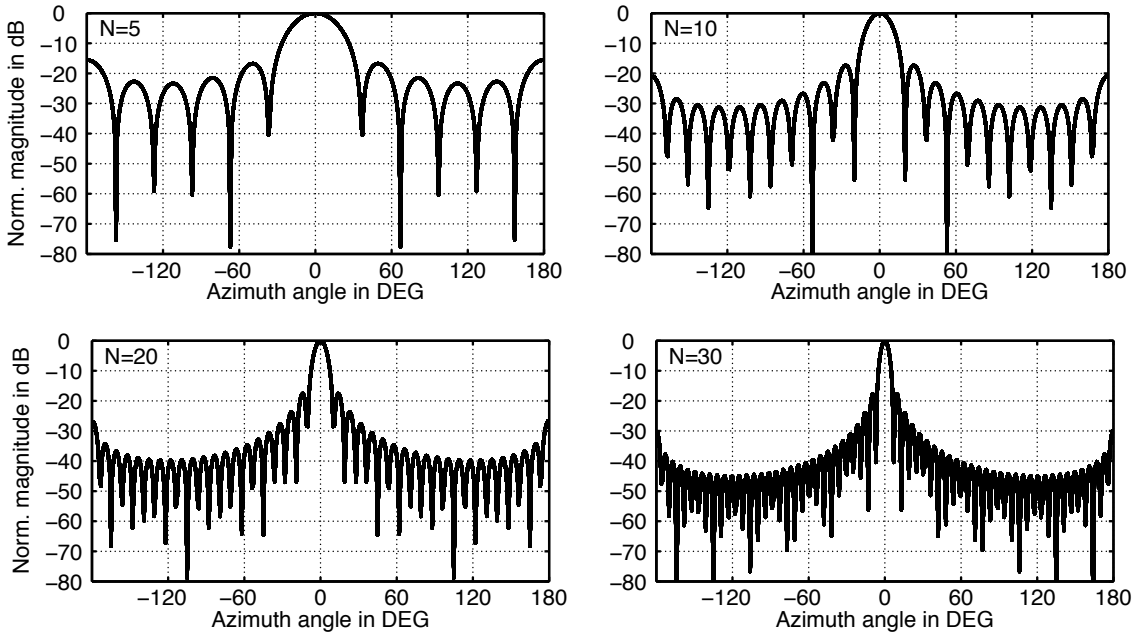

Figure 2.2: Normalized magnitudes of PW coefficients with respect to the azimuth directions for SH rendering orders $N = 5, N = 10, N = 20, N = 30$. The figure is taken from Bernschütz (2016, p. 73) to illustrate the influence of SH order truncation. Smaller N leads to less side-lobes with higher magnitudes, larger N to an increasing number of side-lobes, but with decreasing magnitudes. An ideal PWD with an infinite SH order leads to an ideal spatial Dirac impulse.

2.3 Technical Constraints

2.3.1 Spatial Aliasing

Besides other constraints, as the gain limitation of the radial filters for lower frequencies to avoid noise amplification, the major problem of real-world microphone arrays is the limited microphone distribution density that leads to spatial undersampling. One consequence of those undersampled sound field captures is named spatial aliasing and will be discussed in this section.
Sampling a continuous-time signal basically leads to a replication of the corresponding spectrum at half the sampling frequency (Nyquist-Frequency). When sampling slower than twice of the highest signal frequency, frequency components of the spectral replications are aliased into the lower frequency bands (Shannon, 1998). This temporal aliasing effect is illustrated in Figure 2.3. Same applies when spatial sampling a space-continuous sound field. Sampling with a limited number of sampling points results in the replication of the modal spectra in the wave spectrum domain. Those replications are aliased into the lower modal wave spectra, as can be

seen in Figure 2.4. However, for spatial aliasing, there is no certain temporal frequency bound, as the Nyquist-Frequency for the continuous-time sampling, where no aliasing artifacts arise. Natural sound fields are not band-limited, and thus spatial aliasing is present for every temporal frequency. Nevertheless, higher temporal frequencies are more affected by spatial aliasing than lower temporal frequencies. According to Rafaely (2005), a temporal frequency f_a can be estimated where spatial aliasing artifacts increase rapidly

$$f_A = \frac{N_{sg} c}{2\pi r_0}. \tag{2.21}$$

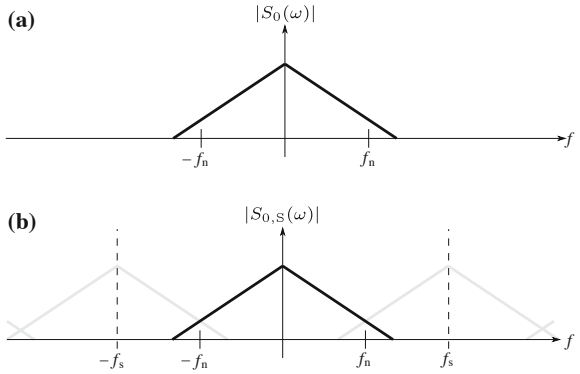

Figure 2.3: Sampling a time continuous signal with a sampling rate smaller than twice of the highest signal frequency leads to the time-aliasing effect. Components of the spectral replications are aliased into the higher frequencies. The figure is taken from Ahrens (2012, p. 119).

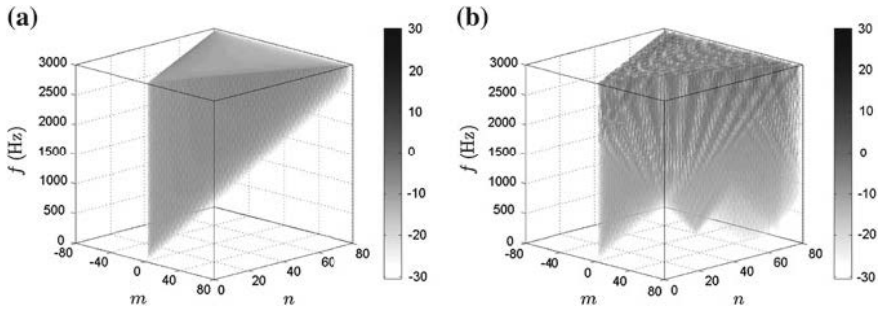

Figure 2.4: Sampling a space-continuous signal results in replications in the wave spectrum domain. Analogous to the time-aliasing, portions of the wave spectrum replications of higher modes are aliased into lower modes. The figure is taken from Ahrens (2012, p. 125).

Thus, SH coefficients of orders $N < N_{sg}$ are not mentionably affected by spatial aliasing, or in other words, to avoid spatial aliasing up to certain temporal frequency f, a grid order of $N_{sg} > (2\pi f r_0)/c = k\ r_0$ is necessary. This frequency f_a will be denoted as spatial alias frequency in the following.
According to Ben-Hur et al. (2018), the aliasing error can be mathematically expressed with

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\underbrace{Y_n^m(\theta',\phi')^* Y_n^m(\theta_d,\phi_d)}_{\{\delta_{nn'}\delta_{mm'}+\epsilon(n,m,n',m')\}}. \tag{2.22}$$

The SH basis functions are not completely orthogonal and thus produce an additional error ϵ.
Neglecting trivial solutions as increasing the number of microphones, or decreasing the array radius r_0 to improve the microphone density, there exist a small number of approaches to reduce the impact of spatial aliasing. These approaches are discussed in Section 2.3.4

2.3.2 Truncation Error

A further essential constraint induced by sparse microphone arrangements is the SH series truncation error. When describing HRIRs or spatial sound fields using SH coefficients, in real-life implementations just a limited modal order of SH basis functions can be used. The DSFT is limited to M_{sg}, as can bee seen in Equation 2.19. However, HRIRs and natural sound fields are not limited in their spatial resolution and thus would require an SH representation with infinite modal order. It is unavoidable to truncate the natural SH order series at a certain point and cut away the higher SH modes that contain important spatial details.
This is highly relevant, especially for HRIRs. The nature of the human pinna induces relevant localization cues at high-frequency components. Representing this localization information in the SH domain requires a high spatial resolution and hence higher SH orders. Neglecting these modes leads to audible artifacts in binaural renderings, as less spaciousness or coloration effects. In particular, the contralateral ear signal that is not exposed by direct sound incidences is more affected by the truncation error.
HRIR data sets are usually measured on sampling grids with an adequate resolution to store all necessary spatial details. However, there are just a limited number of sound field signals obtained by spherical microphone arrays. When merging high modal resolution HRIRs and limited modal resolution sound field data, the truncation effect, and thus loss of spatial detail is unavoidable.
Furthermore, similar to the windowing of a time-signal with a harsh rectangular window which results in side-lobes in the frequency domain, harsh truncating of the SH order results in artifacts in the wave spectrum domain. This effect is illustrated in Figure 2.2 or Figure 2.5. Truncating the SH order series at lower orders leads to fewer side-lobes but with higher amplitude and energy, truncating at higher orders yields multiple side-lobes with decreased magnitudes. Figure 2.5 (bottom, right) was generated by rendering up to an SH order $N = 85$ to simulate a nearly ideal spatial Dirac impulse.

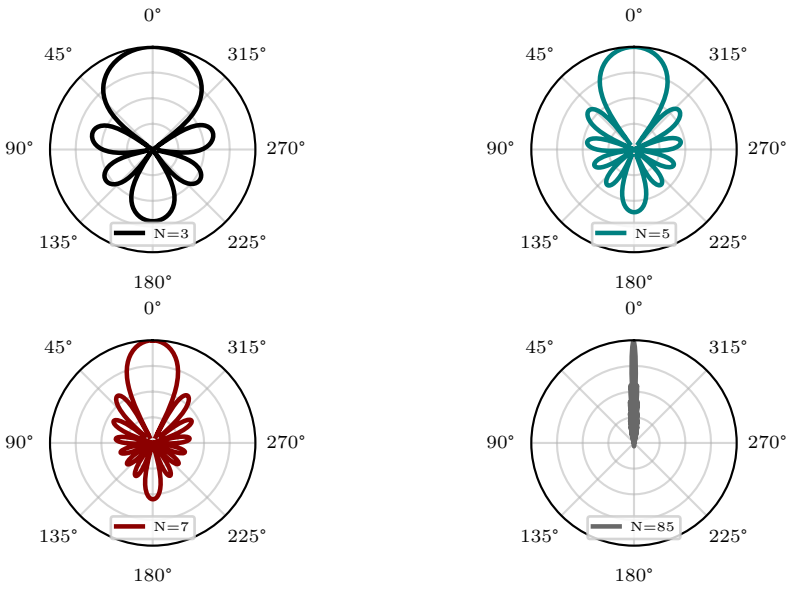

Figure 2.5: Normalized logarithmic magnitudes of SH coefficients of a 0° spatial Dirac impulse computed with Equation 2.8 for maximum SH orders $N = 3$ (black) $N = 5$ (blue) $N = 7$ (red), and $N = 85$ (grey).

2.3.3 Consequences of Undersampled Microphone Array Data Renderings

Using undersampled microphone array recordings for rendering a VAE results in the spatial aliasing effect and at the same time to the SH order truncation. Therefore, it is hard to distinguish between the consequences of those effects. However, spatial aliasing solely depends on the number of microphones, or more precisely on the density of the microphone arrangement, whereas the truncation effect just depends on the SH rendering order. Thus, both effects result in different and even contrary audible artifacts. Ben-Hur et al. (2019) denoted the overall error caused by sparse microphone constellations as sparsity error and mathematically break down the differences of the aliasing and truncation effect.
Spatial undersampling yields two different artifacts in binaural reproductions. On the one hand, the spectral modification, and on the other hand, a loss of spatial information. Figure 2.6 illustrates both of the artifacts. Similar to Figure 2.1, it shows the magnitudes of the PW coefficients of a simulated broadband plane wave impinging on an array at $\phi = 0°, \theta = 90°$ with respect to the frequency and azimuth angle. This time, on a real-life array with 10 cm radius and a 5th order Lebedev sampling scheme. Whereas for the ideal array the magnitudes for certain azimuth directions keep constant for all frequencies, the spatial resolution for the real-life array is totally lost for high frequencies. Additionally, at high frequencies an overall increased level of the PW coefficients can be observed.

The influence of spatial aliasing and truncation error on the time-frequency spectrum is illustrated in Figure 2.7. Both diagrams show the frequency responses of 90° BRIRs to the left ear resulting from binaural reproductions of array impulse re-

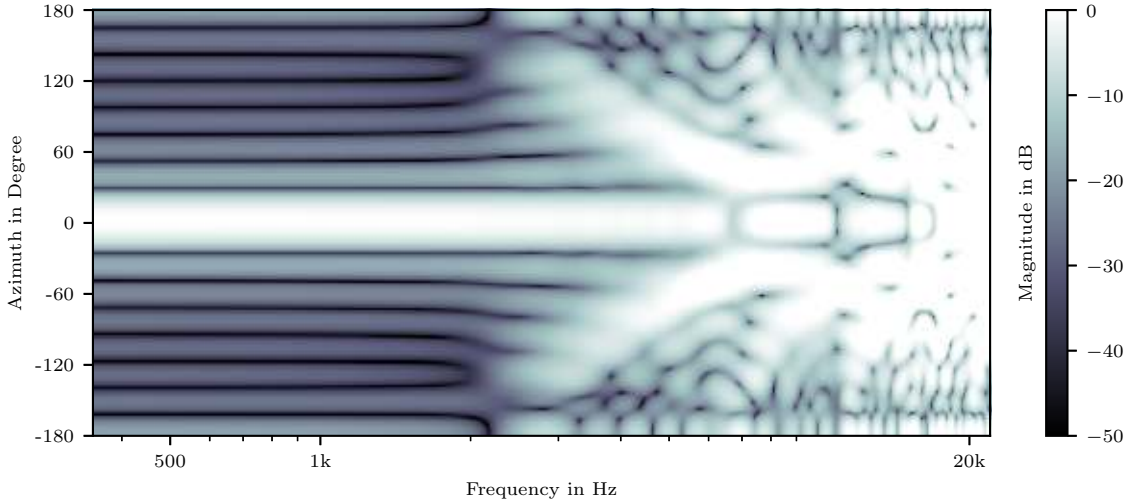

Figure 2.6: Plane wave impact at $\phi = 0°$, $\theta = 90°$ on a real-life microphone array with a radius of 10 cm and a 7th order Lebedev. The figure shows the magnitudes of the PW coefficients for different azimuth angles at the array surface with respect to the frequency. The PW coefficients was rendered up to an SH order of 7.

sponses measured in Control Room 7 at the WDR broadcasting studios (Section 3.1 gives a more detailed explanation of the array measurements by Stade et al. (2012) used for the renderings in this work). The left-hand figure depicts BRTFs (binaural room transfer function) based on 1202 sampling point Lebedev grid array impulse responses. The black curve BRTF was rendered up to an SH order of 29, the blue one up to an SH order of 5. Hence, both BRTFs are not noteably affected by spatial aliasing, but the blue curve is significantly more affected by the truncation error. It can be seen that the truncation error yields a decreased level for higher frequencies, and thus has a low-pass characteristic. The right-hand diagram shows BRTFs based on a 1202 Lebedev grid (black curve) and a 50 sampling point Lebedev grid (blue curve), both rendered up to an SH order of 5. Hence, both renderings contain the same truncation error, but just the blue BRTF is influenced by spatial aliasing. It can be observed that spatial aliasing leads to an increased level at higher frequencies, and thus has a high-pass characteristic. Both illustrations show the effect on the time-frequency spectrum for one direction only. However, it can be shown that spatial aliasing is present with the same amount for all directions, whereas the truncation error is highly direction dependent and yields spectral artifacts in particular for lateral directions, see e.g. Zaunschirm et al. (2018), or Hold et al. (2019).

2.3.4 Reducing the Impairment of Spatial Undersampling

This section gives an overview of common approaches to reduce the perceptional artifacts of spatial undersampled array auralizations. It will be separated between aliasing and truncation mitigation approaches, however some of them handle both of the effects.

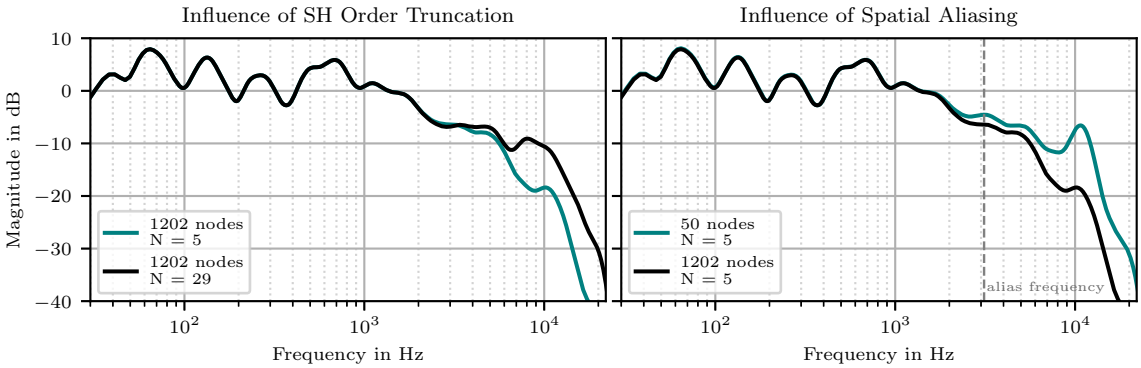

Figure 2.7: 90° BRTF renderings to the left ear based on array impulse responses in Control Room 7 at the WDR studios. The left-hand figure displays two BRTFs resulting from measurements on a 1202 sampling point Lebedev grid, the black one rendered up to a maximum SH order of 29, the blue one up to an SH order of 5. The differences at higher frequencies illustrate the spectral influence of the truncation error. The right side figure shows two BRTFs resulting from renderings up to an SH order of 5. The blue curve resulting from a measurement on a Lebedev grid with 50 sampling points, the black one from a Lebedev grid with 1202 sampling points. This illustrates the spectral influence of spatial aliasing. Both BRTFs were fractional-octave smoothed over 1/3 octave (This holds for every figure displaying frequency responses in this work).

2.3.4.1 Spectral Equalization

As already mentioned, one artifact of spatial aliasing is the increased sound pressure level for higher frequencies. The simplest way to compensate for this spectral artifact is to equalize the resulting binaural signals. Bernschütz (2016, pp.129-130) found that the spectral increase follows a certain regularity. Figure 2.8 (left) taken from Bernschütz (2016, pp.130) shows the logarithmic deviation of 7th order array renderings and the corresponding dummy head measurements, averaged over 360 azimuth directions. For the array rendering, the authors used HRTFs with a limited modal resolution (RHRTFs) (Bernschütz, 2016, pp.83-86). These RHRTFs are based on SH interpolation and prevent the truncation effect, whereas the depicted frequency responses are just affected by spatial aliasing. Figure 2.8 illustrates that for diffuse sound fields, the deviation nearly linear increases with a 6 dB per octave slope for frequencies above the spatial alias frequency. The right-hand figure shows the resulting compensation filter, generated by inverting the average deviation curve. To calculate a generative filter with respect to the array configuration, the author proposed an ordinary first-order low-pass filter with a cut-off at the spatial alias frequency.

To evaluate this, Figure 2.9 was created. It shows the average logarithmic deviation of binaural signals, rendered with 1202 and 50 sampling point array impulse responses on the same SH order. Just like in Figure 2.7 (right), the 50 sampling point rendering is affected by spatial aliasing. The blue curve in the upper figure shows the logarithmic deviation of a 90° BRTF to the left ear, the black curve the logarithmic deviation averaged over right and left BRTFs for 360 azimuth directions. Bernschütz (2016) compared dummy head BRIRs to RHRTF array renderings and

found that the deviation follows a 6 dB per octave slope. However, in Figure 2.9, it can be observed that the deviation increases nearly linear with 12 dB per octave. The generative low-pass filter matching the inverse of this curve can be calculated by second order low-pass filters, depicted in the bottom figure for array orders 5, 7 and 10. Nevertheless, for consistency, the discussion and preparation of the listening experiment stimuli are referred to the first-order equalization filters as proposed by Bernschütz (2016).

Figure 2.10 illustrates the result of the equalization, using second order low-pass filters. Again frequency responses of BRIRs rendered with 1202 and 50 sampling point array data are compared. Additionally, the blue curve shows the equalized curve which closely matches the desired grey 1202 sampling point BRTF.

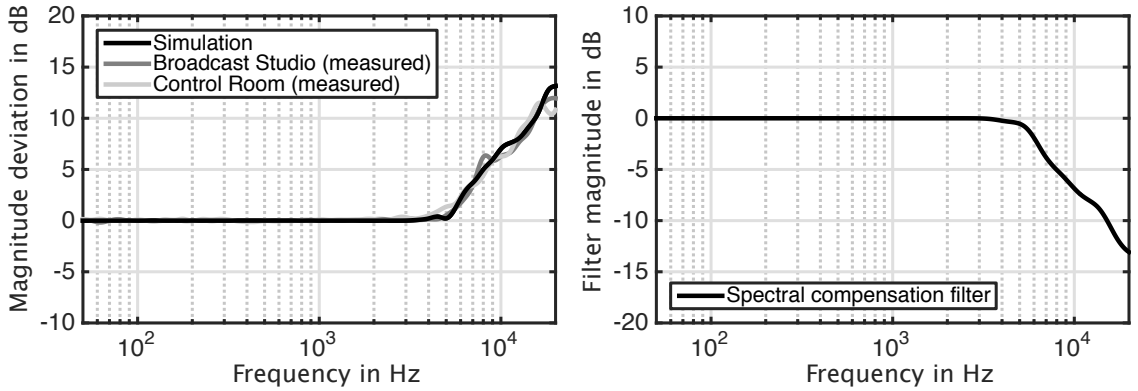

Figure 2.8: On the left-hand side the average logarithmic deviations of 7th order array renderings and corresponding dummy head measurements are depicted. The deviation matches a linear 6 dB per octave increase for frequencies above the alias frequency. The right-hand diagram illustrates the resulting compensation filter derived from inverting the average deviation curve. The picture taken from Bernschütz (2016, pp.130)

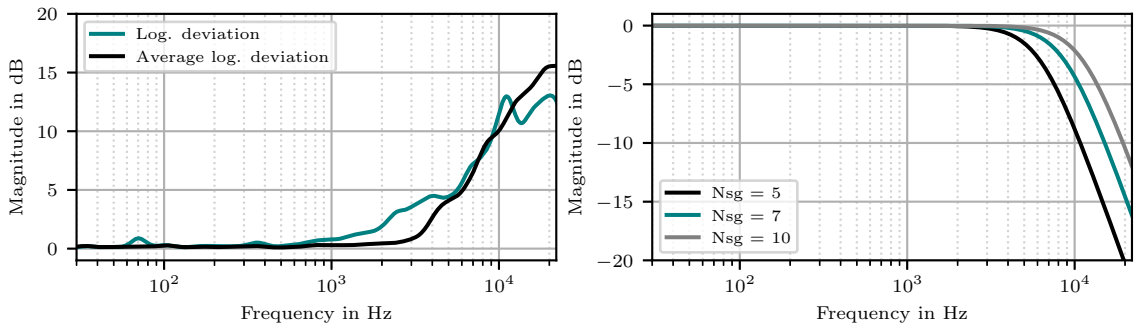

Figure 2.9: Logarithmic deviation curve of 5th order array renderings done with 50 and 1202 sampling point array measurements. The blue curve depicts the logarithmic deviation for a 90° BRTF to the left ear, the black curve the deviation averaged over both ears and 360 azimuth directions. The bottom figure displays generative second order low-pass filters for the SH order 5, 7 and 10.

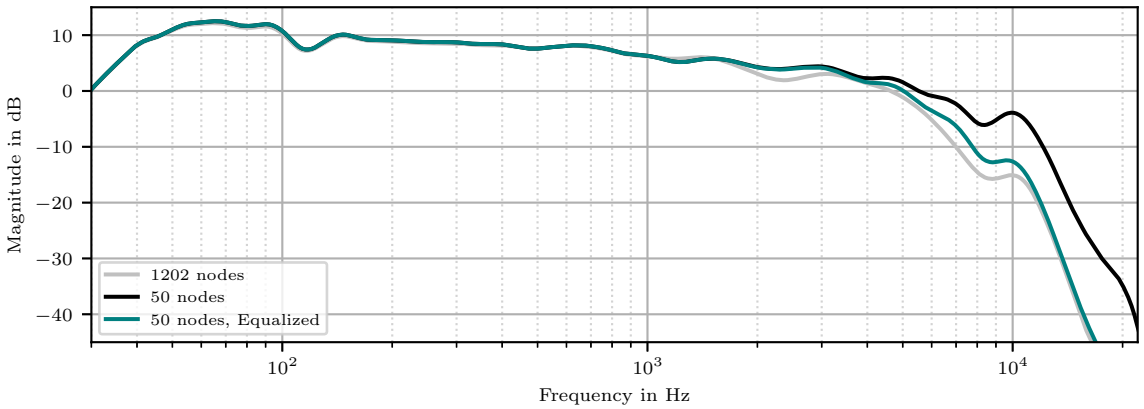

Figure 2.10: 90° BRTF to the left ear based on array renderings of 1202 (grey) and 50 (black and blue) sampling point Lebedev grid measurements. Both were rendered up to an SH order of 5. The black curve is significantly affected by spatial aliasing, the blue curve was equalized by a second order low-pass filter.

Similarily, the spectral modification caused by the truncation error can be equalized by appropriate compensation filters. Ben-Hur et al. (2017) introduced so-called Spherical Head Filters (SHF) which also can be applied to the binaural signals. Similar to the assumptions for the spatial aliasing compensation filters, Ben-Hur et al. (2017) expect an average response of a diffuse sound field, but additionally assume a rigid sphere approximation for the human head. Figure 2.11 depicts the proposed SHF for different SH orders N. These filters were designed under the assumption of negligible spatial aliasing.

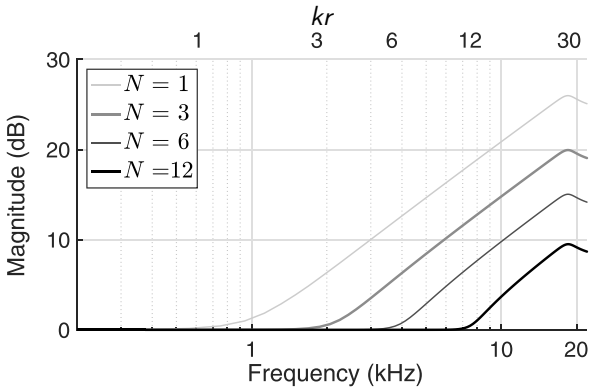

Figure 2.11: The so-called Spherical Head Filters for SH orders 1, 3, 6 und 12 introduced by Ben-Hur et al. (2017). They assume negligible spatial aliasing and compensate for spectral artifacts induced by the SH order truncation.

Both spectral compensation filters are either proposed as truncation compensation or spatial aliasing compensation filters. However, in real-life both effects are present at the same time. In Section 2.3.5.5 a combination of both filters compensating the overall undersampling error is discussed.

2.3.4.2 Bandwidth Extension Algorithm for Microphone Arrays

The Bandwidth Extension Algorithm for Microphone Arrays (BEMA) has been developed as spatial aliasing mitigation approach, see Bernschütz (2012) or Bernschütz (2016). The sound field SH coefficients of higher frequency bands that are affected by spatial aliasing, will be substituted with synthesized SH coefficients. As all higher SH coefficients will be replaced, the BEMA approach compensates not just the spatial aliasing artifacts, but also the truncation error. Basic idea is to use certain properties of the reliably obtainable frequency bands to estimate the SH coefficients of the higher frequencies. Therefore, spatial and spectral properties will be treated separately. The spatial energy distribution will be extracted from the SH coefficients of frequency bands below the aliasing frequency, captured by the microphone array. The total energy of the higher frequencies will be obtained by an additional omnidirectional microphone, ideally located in the array center. This approach is based on the assumption of a natural spatial sound field, where the energy distribution of adjacent frequency bands follows a certain similarity.

The sound field SH coefficients $\mathring{\mathbf{S}}_{nm}(\omega)$ with respect to the frequencies $(\omega_0, \omega_1, \ldots, \omega_K)$ can be expressed as the $(N+1)^2 \times (K+1)^2$ matrix

$$\mathring{\mathbf{S}}_{nm}(\omega) = \begin{bmatrix} \mathring{S}_{00}(\omega_0) & \mathring{S}_{00}(\omega_1) & \ldots & \mathring{S}_{00}(\omega_K) \\ \mathring{S}_{-11}(\omega_0) & \mathring{S}_{-11}(\omega_1) & \ldots & \mathring{S}_{-11}(\omega_K) \\ \mathring{S}_{01}(\omega_0) & \mathring{S}_{01}(\omega_1) & \ldots & \mathring{S}_{01}(\omega_K) \\ \mathring{S}_{11}(\omega_0) & \mathring{S}_{11}(\omega_1) & \ldots & \mathring{S}_{11}(\omega_K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathring{S}_{NN}(\omega_0) & \mathring{S}_{NN}(\omega_1) & \ldots & \mathring{S}_{NN}(\omega_K) \end{bmatrix}. \tag{2.23}$$

The matrix $\mathring{\mathbf{S}}^a_{nm}(\omega)$ (Eq. 2.24) as a subset of $\mathring{\mathbf{S}}_{nm}(\omega)$ with respect to the frequencies $(\omega_a, \ldots, \omega_{K-1}, \omega_K)$ holds the aliased SH coefficients which will be replaced with the synthesized BEMA SH coefficients. ω_a can be approximated with $\omega_a \approx \frac{N_{sg}}{r_0}$ according to 2.21.

$$\mathring{\mathbf{S}}^a_{nm}(\omega) = \begin{bmatrix} \mathring{S}_{00}(\omega_a) & \ldots & \mathring{S}_{00}(\omega_{K-1}) & \mathring{S}_{00}(\omega_K) \\ \mathring{S}_{-11}(\omega_a) & \ldots & \mathring{S}_{-11}(\omega_{K-1}) & \mathring{S}_{-11}(\omega_K) \\ \mathring{S}_{01}(\omega_a) & \ldots & \mathring{S}_{01}(\omega_{K-1}) & \mathring{S}_{01}(\omega_K) \\ \mathring{S}_{11}(\omega_a) & \ldots & \mathring{S}_{11}(\omega_{K-1}) & \mathring{S}_{11}(\omega_K) \\ \vdots & \ddots & \vdots & \vdots \\ \mathring{S}_{NN}(\omega_a) & \ldots & \mathring{S}_{NN}(\omega_{K-1}) & \mathring{S}_{NN}(\omega_K) \end{bmatrix} \tag{2.24}$$

To extract the reliably spatial information denoted as spatiotemporal image $\mathring{I}_{mn}$ an appropriate source band has to be defined. It is preferable to choose source bands as close as possible to the synthetic bands. In the following, the bands will be assumed as $\omega_a - \mu$. The spatial image can be computed with

$$\mathring{I}_{nm} = \frac{1}{W} \sum_{\mu=1}^{W} d_n \left(\frac{\omega_a - \mu}{c} r_0 \right) \mathring{\mathbf{S}}_{nm}(\omega_a - \mu). \tag{2.25}$$

By multiplying with d_n, the dependency on the radial filter vanishes. The spatiotemporal image should just hold information about the energy distribution, as the total energy will be taken from the center microphone. Therefore, it has to be power normalized according to

$$\overline{p} = \frac{1}{W} \sum_{n=0}^{N} \sum_{m=-n}^{m} \sum_{\mu=1}^{W} \left| d_n \left(\frac{\omega_a - \mu}{c} r_0 \right) \mathring{\mathbf{S}}_{nm}(\omega_a - \mu) \right|^2 \tag{2.26}$$

$$p = \sum_{n=0}^{N} \sum_{m=-n}^{m} \left| \mathring{I}_{mn} \right|^2 \tag{2.27}$$

$$\mathring{I}_{nm}^{\text{norm}} = \sqrt{\frac{\overline{p}}{p}}\, \mathring{I}_{nm}. \tag{2.28}$$

The center signal will be normalized to its own average level in the source band $(\omega_a - \mu)$ and the synthetic BEMA SH coefficients can be calculated with

$$\mathring{\mathbf{S}}_{nm}^{\text{BEMA}} = \underbrace{\frac{1}{d_n(\frac{\omega}{c} r_0)} \mathring{I}_{nm}^{\text{norm}}}_{\text{spatial information}} \underbrace{c^{\text{norm}}(\omega)}_{\text{spectral information}} \quad . \tag{2.29}$$

Here, the corresponding radial filter d_n will be taken into account. The authors propose to optimize the phase of the synthetic SH coefficients. For this, the phase of the synthetic SH coefficients will be shifted according to a phase offset $\Delta\phi$ obtained at the transition bin

$$\Delta\phi = \arg \left(\frac{\mathring{S}_{00}(\omega_a)}{\mathring{S}_{00}^{\text{BEMA}}(\omega_a)} \right) \tag{2.30}$$

$$\overline{\mathring{\mathbf{S}}}_{nm}^{\text{BEMA}}(\omega) = \mathring{\mathbf{S}}_{nm}^{\text{BEMA}}(\omega)\, \mathrm{e}^{j\Delta\phi}. \tag{2.31}$$

Further, it is reasonable to fade in the synthetic coefficients, instead of a harsh crossover. Bernschütz (2012) introduced one feasible crossover method based on a smooth overlapping of synthetic and original SH coefficients. This will not be derived in more detail.
For a single plane wave, the knowledge of the pressure distribution on the array surface for a single frequency bin allows to obtain the pressure distribution over the entire frequency range. Hence, BEMA works perfectly for a single plane wave. However, Bernschütz (2016, pp.144-146) found out that even for three time shifted plane waves impinging on the array surface the BEMA algorithm result in notable artifacts. It is unclear, if BEMA improves the auralization of array recordings in diffuse sound fields.

2.3.4.3 Magnitude Least Squares

Schörkhuber et al. (2018) introduced a method for reducing the impact of order truncation denoted as Magnitude Least-Squares (MagLS). This method optimizes the used HRTF set in such a way that they will lead to lower truncation errors after

the binaural rendering. The MagLS approach consists of to parts. First of all, the HRTFs will be modified such that the energy in higher SH orders is reduced without noteably decreasing the perceptional quality.
This modification is an advancement of the time alignment (TA) approach by Zaunschirm et al. (2018). According to the Duplex Theory by Rayleigh (1907), for high frequencies the Interaural Time Differences (ITDs) become perceptional less important than the Interaural Level Differences (ILDs). However, most of the energy in higher modes is caused by rapid phase changes towards higher frequencies. Thus, removing the linear phase at higher frequencies will decrease the energy in higher modes without significantly modifying the ILDs

$$\mathbf{H}_{\mathrm{TA}}(\Omega,\omega) = \begin{cases} \mathbf{H}(\Omega,\omega) & \text{if } \omega \leq \omega_c \\ \mathbf{H}(\Omega,\omega)\mathrm{e}^{-\mathrm{j}\phi_l(\Omega,\omega)} & \text{if } \omega \geq \omega_c \end{cases}. \tag{2.32}$$

Where $\phi_l(\Omega,\omega)$ is the linear phase with respect to the ITD for direction Ω and ω_c the cut-off frequency which was empirically chosen by the authors. MagLS aims at finding the optimum phase instead of completely removing the linear phase. This can be achieved by solving the least-square problem that minimizes the distance of the magnitudes of the modified HRTFs and a reference set

$$\min_{\mathring{\mathbf{H}}_{nm}(\omega)} \sum_{sp=1}^{M_{sg}} [|Y_n^m(\Omega_{sg})\mathring{\mathbf{H}}_{nm}(\omega)| - \mathbf{H}(\Omega_{sg},\omega)]^2. \tag{2.33}$$

It would be beyond the scope of this thesis to derive this least-square procedure why the reader is referred to Schörkhuber et al. (2018).

After modifying the phase for higher frequencies either with the TA or the MagLS approach, in a second step, the HRTFs will be optimized using a so-called covariance constrained matrix $\mathbf{R}(\omega)$. This matrix will equalize the spectrum based on a statistical diffuse sound field expectation. For a detailed derivation of $\mathbf{R}(\omega)$ the reader is referred to Zaunschirm et al. (2018), or Zotter and Frank (2019).

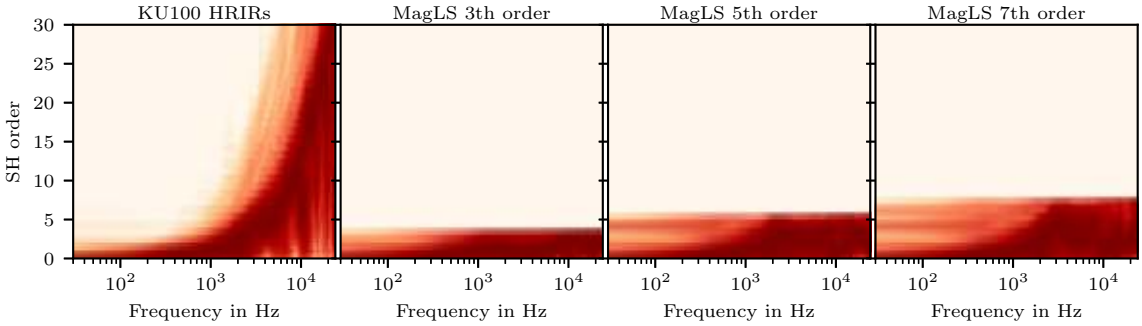

Figure 2.12: Normalized logarithmic energy distributions of HRTF SH coefficients rendered up to the orders $N = 1$ to $N = 30$. All energy plots are based on a KU100 HRIR dataset. The left-hand figure depicts the energy distribution of the untreated HRIR set, the right-hand figures the distributions of the MagLS pre-processed HRIRs for target rendering orders $N = 3$, $N = 5$, and $N = 7$. The color encodes the energy ranging from -40 dB to 0 dB normalized to the maximum values for each frequency bin.

Figure 2.12 illustrates the efficient energy concentration of the MagLS approach for the different target rendering orders $N = 3, N = 5$, and $N = 7$. All four plots are based on the KU100 2707 HRIR dataset by Bernschütz (2013). They are displaying the energies of the SH coefficients rendered for orders $N = 1$ up to $N = 30$ with respect to the frequency, normalized to the maximum value for each frequency bin. The left-hand figure was processed with the untreated HRIR set, the right-hand figures for HRIRs that were pre-processed for target SH rendering order 3, 5 and 7. The color indicates the energies ranging from -40 dB to 0 dB. It can clearly be seen that for the pre-processed HRIR sets the energy is concentrated to the lower orders.
There are a number of further HRIR pre-processing approaches that have been discussed and evaluated by Brinkmann and Weinzierl (2018).

2.3.4.4 Tapering

Hold et al. (2019) introduced a method denoted as spherical harmonic tapering to reduce the side-lobes induced by order truncation. Hard truncating the SH order is equivalent to window the SH coefficients with a rectangular window which results in unwanted side-lobes. Windowing the SH coefficients using appropriate smoother window functions yields a fade-out of higher order modes which ideally results in side-lobe supression. Hold et al. (2019) proposed a half-sided Hanning window to yield this softer truncation. Furthermore, the authors propose to equalize the binaural signals using the Spherical Head Filters discussed in the previous section. However, the Hanning tapering requires slightly modified SHF. These modified filters are depicted in Figure 2.14.

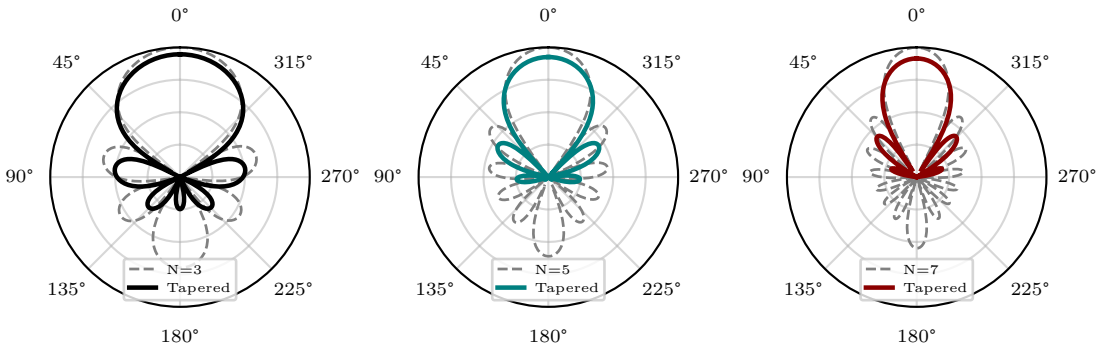

Figure 2.13: Normalized logarithmic magnitudes of SH coefficients of a spatial Dirac impulse in 0° direction computed with Equation 2.8 for maximum SH orders $N = 3$, $N = 5$, and $N = 7$. For each order the rectangular windowed (grey dashed) and Hanning windowed (colored) SH coefficients are depicted. SH Tapering using a Hanning window significantly results in side-lobe suppression, but also in a wider main-lobe with an increased level.

2.3.4.5 Matrix Regularization

Alon and Rafaely (2012) (Alon and Rafaely, 2017) developed a method for spatial aliasing cancellation based on a matrix regularization. For this, it is advantageous to introduce the matrix denotation of the PWD:

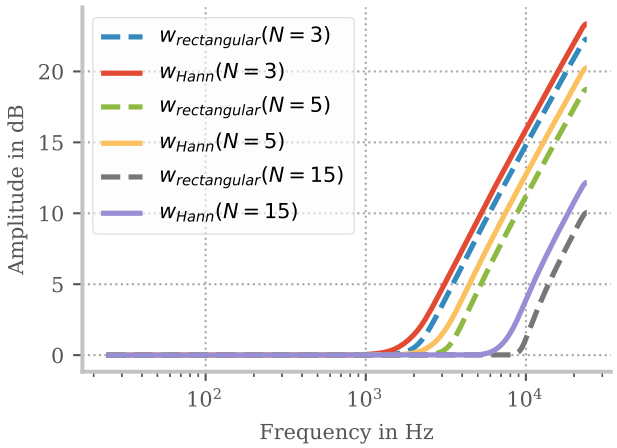

Figure 2.14: Hold et al. (2019) propose to equalize the resulting binaural signals after weighting the sound field SH coefficients with a tapering window. The Spherical Head Filters have to be adapted with respect to the window which is shown in the figure, taken from Hold et al. (2019). The adaptation basically results in a shift of the cut-off towards higher frequencies.

According to Equation 2.15, the PW components $D(\omega)$ can be calculated from the sound pressure SH coefficients $\mathring{\mathbf{P}}_{nm}$ as follows

$$\mathbf{D} = \mathbf{Y}\,\mathrm{diag}(b)\mathring{\mathbf{P}}_{nm}. \tag{2.34}$$

With the radial functions b and the SH basis functions Y

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\Omega_1) & Y_1^{-1}(\Omega_1) & \dots & Y_{N_{sg}}^{N_{sg}}(\Omega_1) \\ Y_0^0(\Omega_2) & Y_1^{-1}(\Omega_2) & \dots & Y_{N_{sg}}^{N_{sg}}(\Omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\Omega_{M_{sg}}) & Y_1^{-1}(\Omega_{Msg}) & \dots & Y_{N_{sg}}^{Nsg}(\Omega_{Msg}) \end{bmatrix}. \tag{2.35}$$

As long as the sound field is limited to an SH order of N, thus $M_{sg} \geq (N+1)^2$ is fulfilled, the computation of the PW components from the sound pressure signals $P(\Omega, \omega)$ can be assumed to be error free. By rearranging this equation, an expression for the sound field SH coefficients can be found

$$\mathring{\mathbf{P}}_{nm} = \mathrm{diag}(b)^{-1}\mathbf{Y}^*\mathbf{D}. \tag{2.36}$$

However, real-life sound fields are not limited in their expansion order. Hence, for undersampled sound field recordings with $M_{sg} \leq (N+1)^2$ the PWD produces an error and Equation 2.3.4.5 transforms to

$$\mathbf{D} = \widetilde{\mathbf{Y}}\mathrm{diag}(\tilde{b})\widetilde{\mathring{\mathbf{P}}}_{nm}. \tag{2.37}$$

With $\widetilde{\mathbf{Y}} = \left[\mathbf{Y}, \mathbf{Y}_\Delta\right]$, $\tilde{b} = \left[b^T, b_\Delta^T\right]^T$, and $\widetilde{\mathring{\mathbf{P}}}_{nm} = \left[\mathring{\mathbf{P}}_{nm}^T, \mathring{\mathbf{P}}_\Delta^T\right]^T$. Where $(\cdot)_\Delta$ denotes the spatial aliasing affected higher order parts. Consequently, the aliased SH coefficients can be expressed as

$$\mathring{\mathbf{P}}_{nm}^{\text{alias.}} = \text{diag}(b)^{-1}\mathbf{Y}^{*}\mathbf{D} = \underbrace{\underbrace{\text{diag}(b)^{-1}\mathbf{Y}^{*}}_{\text{stable}}\underbrace{\widetilde{\mathbf{Y}}\text{diag}(\widetilde{b})}_{\text{aliased}}}_{\mathbf{A_p}}\widetilde{\mathring{\mathbf{P}}}_{nm} \tag{2.38}$$

Where the matrix $\mathbf{A_p}$ is denoted as aliasing projection matrix. The first part of $\mathbf{A_p}$ consists of the stable obtainable wave spectrum information, the second part describes the aliased information. It can be written as $\left[\mathbf{1}, \Delta_\epsilon\right]$($\mathbf{1}$ = unit matrix, Δ_ϵ = aliasing error pattern). $\mathring{\mathbf{P}}_{\mathbf{nm}}^{\mathbf{alias.}}$ can thus be broken down to

$$\mathring{\mathbf{P}}_{nm}^{\text{alias.}} = \mathring{\mathbf{P}}_{nm} + \Delta_\epsilon\widetilde{\mathring{\mathbf{P}}}_{\Delta}. \tag{2.39}$$

Aim of the matrix regularization is to minimize the term $\Delta_\epsilon\widetilde{\mathring{\mathbf{P}}}_{\Delta}$. For this purpose, a regularization matrix $\mathbf{C_{AC}}$ is introduced

$$\mathring{\mathbf{P}}_{nm} \stackrel{!}{=} \mathbf{C_{AC}}\,\mathring{\mathbf{P}}_{nm}^{\text{alias.}} = \mathbf{C_{AC}}\,\mathbf{A_p}\,\widetilde{\mathring{\mathbf{P}}}_{nm} = \mathbf{C_{AC}}\mathring{\mathbf{P}}_{nm} + \mathbf{C_{AC}}\,\Delta_\epsilon\widetilde{\mathring{\mathbf{P}}}_{\Delta}. \tag{2.40}$$

One measure to express the overall aliasing error ϵ_{PWD} is the squared Euclidean distance between the ideal SH coefficients $\mathring{\mathbf{P}}$ and the aliased coefficients $\widetilde{\mathring{\mathbf{P}}}$. Thus, ideally, multiplication with $\mathbf{C_{AC}}$ should zero the spatial aliasing error $\epsilon_{PWD} = \left\|\widetilde{\mathring{\mathbf{P}}} - \mathring{\mathbf{P}}\right\|^2$. Substituting $\mathbf{C_{AC}}\,\mathbf{A_p}$ and calculating the minimum by deriving and resolving to zero yields the final aliasing regularization matrix

$$\mathbf{C_{AC}} = (\mathbf{A_p}\,\mathbf{A_p}^{H})^{-1}\mathbf{A_p}. \tag{2.41}$$

With $\mathbf{A_p} = \text{diag}(b)^{-1}\mathbf{Y}^{*}\,\widetilde{\mathbf{Y}}\,\text{diag}(\widetilde{b})$ and $(\cdot)^H$ the Hermetian operator.
Alon and Rafaely (2017) showed that this approach works well for simulated plane wave scenarios. However, it is unclear if it produces satisfying improvements for real-life diffuse sound fields. The matrix regularization approach would be highly suitable for real-time applications, because the C_{AC} matrix just depends on the radial filters and thus could be completely pre-rendered. Furthermore, there are a number of further regularization approaches discussed in literature preventing different artifacts, for example reduction of noise gain. Those regularizations could be easily combined.

2.3.4.6 Spatial Anti-Aliasing Filters

To prevent time aliasing when sampling continuous time signals, low-pass filters are commonly used to reduce the energy in higher frequency components. Thus, it is an obvious approach to apply spatial low-pass filters, also denoted as modal low-pass filters, before spatial sampling. Rafaely et al. (2007) and Meyer et al. (2008) presented a theoretical approach for deploying those modal low-pass filters. Basic idea is to weight the sound pressure function s with a filter function h before spatial sampling in such a way that it contains less energy in higher orders. It was shown

that an ideal filter function that corresponds to a rectangular modal window can be expressed with

$$h(\theta, \phi) = \frac{N+1}{4\pi(\cos\theta - 1)}[P_{N+1}(\cos\theta) - P_N(\cos\theta)]. \tag{2.42}$$

This filter eliminates the energy in all SH modes above N. The spatial filtering process can be described as spherical correlation of the sound pressure s and an additional transducer providing an expanded membrane covering a wider area of the sphere. This additional transducer rotates over the entire surface to weight the incident sound pressure s. Theoretically, this spatial filtering leads to remarkable reductions of spatial aliasing, but has some essential limitations. Firstly, similar to time domain filtering where ideal rectangular windows lead to large side-lobes in the frequency domain, modal rectangular windows yield side-lobes in the spherical wave spectrum domain. To prevent this, Rafaely et al. (2007) and Elahi et al. (2019) discussed different window functions used for spatial filtering which are depicted in Figure 2.15. A further limitation is the realization of an additional expanded transducer itself. The expansion of the sensing membrane could be simulated by multiple narrow microphones, however, the design is highly challenging. Bernschütz (2016, pp.132-138) extensively discusses expanded transducers and mentions an approach for using such a sub-array for implementing the spatial alias filtering.

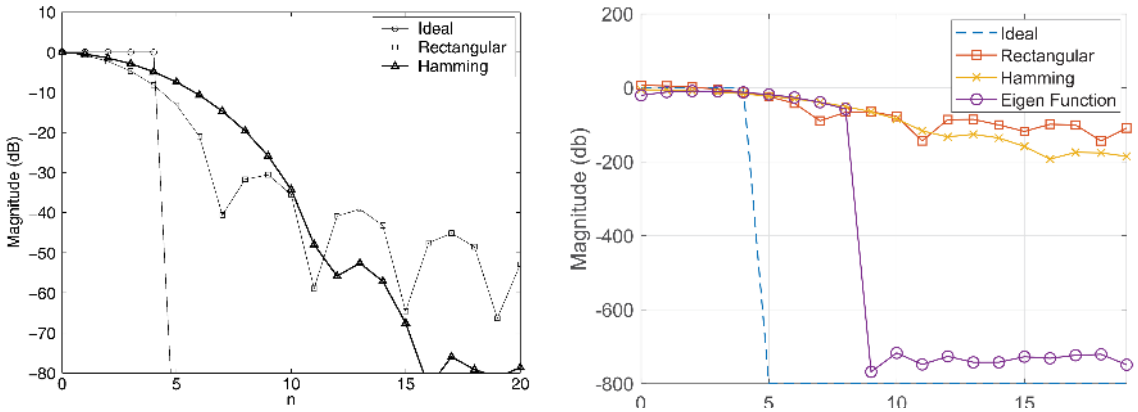

Figure 2.15: The left-hand figure taken from Rafaely et al. (2007) illustrates the magnitudes of the modal-filters designed for reducing the energy of SH orders greater than 4. It shows the ideal spatial aliasing filter, as well as rectangular and Hamming window spatial constrained filters. The right-hand figure taken from Elahi et al. (2019) shows similar filter magnitude responses, but designed for damping SH orders larger than 5. Additionally, the proposed modal-filter based on a slepian eigenfunction window is depicted.

2.3.5 Analysis of the Resulting Binaural Signals

Informal prior listening tests showed that not all approaches described in Section 2.3.4 yield satisfying audible improvements. Thus, just a few algorithms will be investigated in more detail. This section will discuss the most promising approaches and their influence on the binaural signals.

2.3.5.1 MagLS

The Magnitude Least-Squares algorithm (MagLS) is an approach for mitigation of the truncation error. Figure 2.16 and Figure 2.17 show the comparison of two renderings based on array impulse response measurements in the CR7 on a 1202 sampling node Lebedev grid, for the SH orders 29 and 5. This ensures that both renderings are affected by the same spatial aliasing, but by different truncation errors. The figures show left (blue) and right (red) ear signals separately which simplifies to illustrate the directional dependency of the truncation error. In particular, the 90° BRTFs (Fig. 2.17) demonstrate that the contralateral left ear signals are more affected by the truncation error as the ipsilateral right ear signals. The shape of the frequency response to the left ear is totally modified, whereas the right ear signal is just damped with a nearly constant offset towards higher frequencies.
Comparing the 0° and 90° ear signals shows that the truncation error has a different amount of impact for different directions. However, the MagLS algorithm yields a significant improvement for both directions.

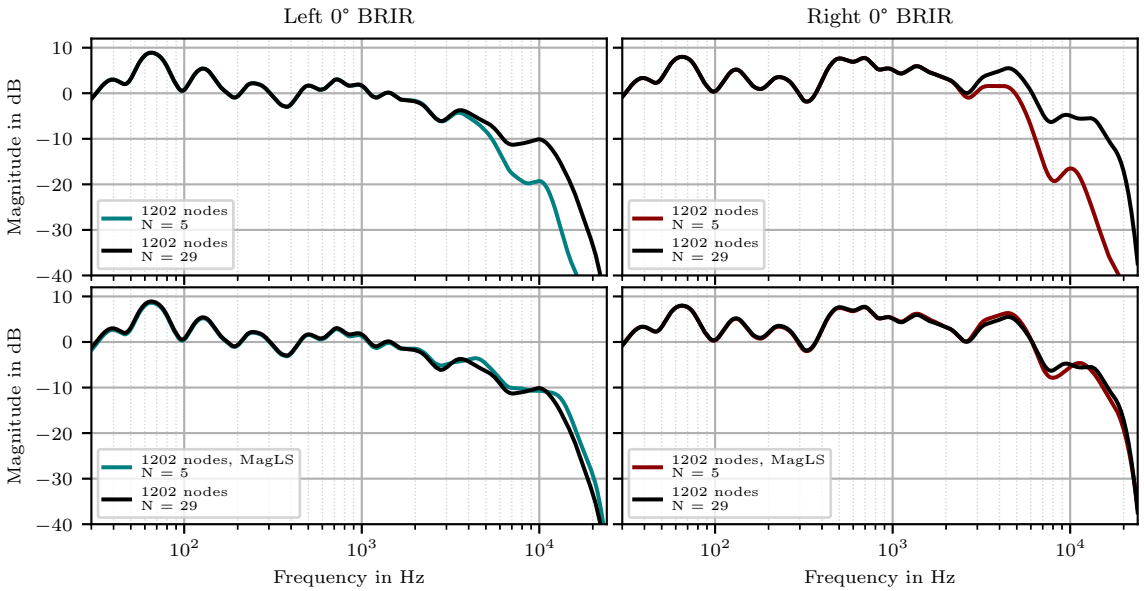

Figure 2.16: MagLS: Frequency responses of 0° left and right ear signals resulting from array renderings based on a 1202 sampling node grid rendered up to an SH order N of 29 (black curves) and 5 (colored curves). The upper diagrams show the influence of the truncation error for the right and left ear separately. The bottom diagrams depict the MagLS improvements. It can be seen that they work equally for both ear signals.

2.3.5.2 Spherical Head Filter - SHF

The Spherical Head Filters (SHF) are global filters applied to the binaural ear signals compensating for the spectral artifacts induced by SH order truncation. Figure 2.18 (left) depicts the SHF filters for the orders $N = 3$, 5 and 7. The improvements that can be achieved with the SHF filtering are illustrated in Figure 2.19. The binaural signals were rendered based on the same array configurations as for the MagLS Figures 2.16 and 2.17, however, just for the left ear. The ear signal for 0° (left) shows a

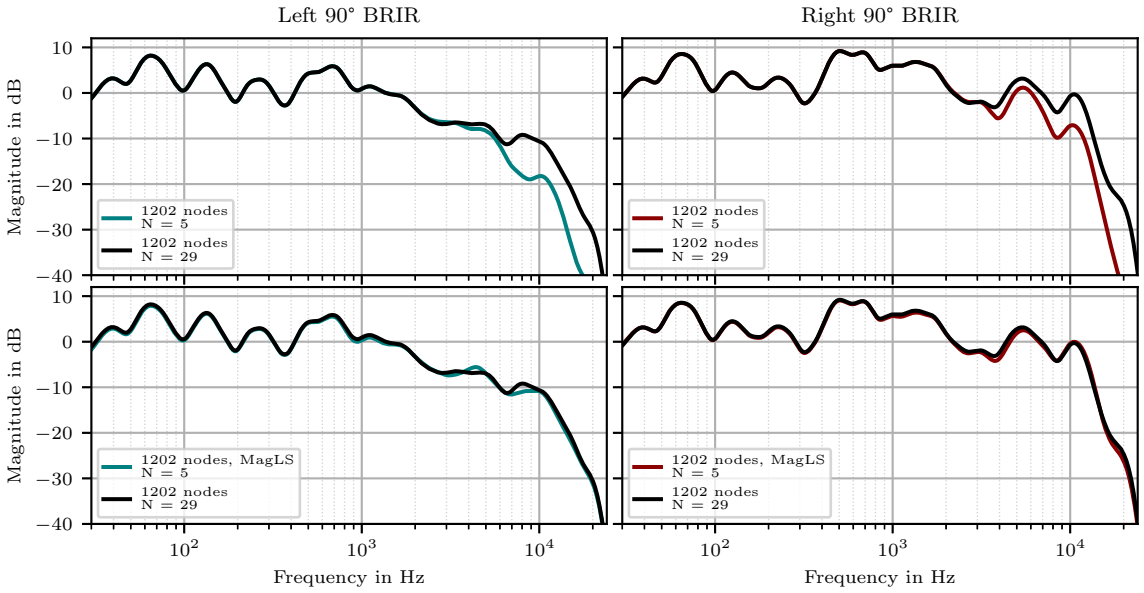

Figure 2.17: MagLS: Frequency responses for 90° left (contralateral) and right (ipsilateral) ear signals resulting from array renderings based on a 1202 sampling node grid rendered up to an SH order N of 29 (black curves) and 5 (colored curves). The upper diagrams show the influence of the truncation error for the right and left ear separately. The bottom diagrams depict the MagLS improvements. Involving the Figure 2.16 shows that MagLS achieves remarkable improvements independent on the direction.

significant improvement of the frequency response. Comparing it to the 90° signals demonstrates that the global filter applied to all directions equally yields unwanted effects for the more critical lateral frequency responses. It thus compensates the coloration for frontal directions more than for lateral directions. However, on average, the SHF filtering leads to a spectral improvement for all directions.

2.3.5.3 Tapering

The Tapering algorithm tries to mitigate the harsh truncation of the SH order series by applying a Hanning window based weighting of the SH coefficients. Additionally, the investigators propose to apply a slightly modified version of the SHF. Figure 2.20 displays the binaural signals resulting from the same renderings as before (e.g. Fig. 2.19), but involving the Tapering algorithm. Since the observation of frequency responses for single directions mainly illustrates the spectral improvements, Figure 2.20 basically indicates the influence of the modified SHF filtering. However, informal listening probes revealed that the Tapering yields audible improvements for lateral directions. The final listening experiment will show if the Tapering has a perceivable influence, or if the SHF filtering yields the more significant improvement.

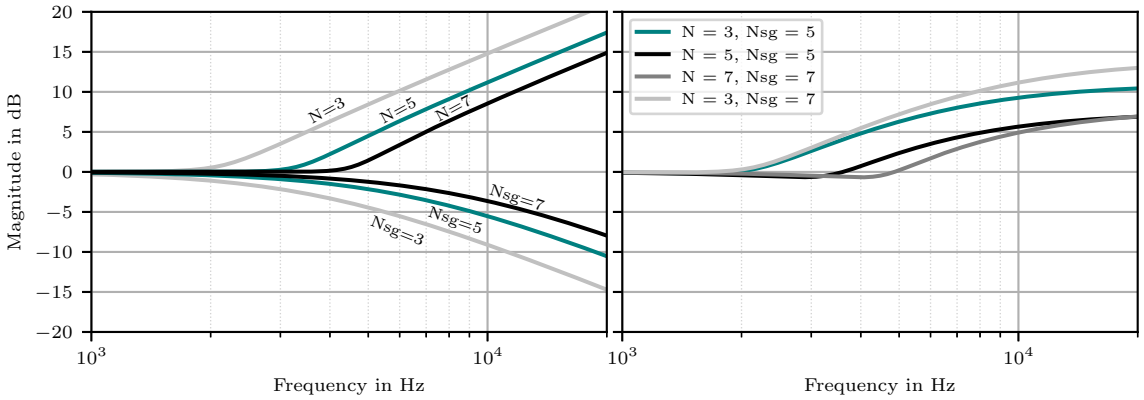

Figure 2.18: SHF: Spherical Head Filters and aliasing compensation filters (left) for multiple SH orders N and grid orders N_{sg}. The aliasing compensation filters (with negative slope) depends on the grid order, the SHF (with positive slope) solely on the SH order. The right-hand figure depicts the combination of both filters for different N-N_{sg} configurations resulting in the global undersampling filters (GEQ).

2.3.5.4 BEMA

Figure 2.21 and Figure 2.22 were generated to study the improvement of BEMA. Therefore, array measurements in the CR7 on a 1202 sampling node grid (black curves) and 50 sampling node grid (colored curves) are rendered up to the same SH order of 5. They are affected by the same truncation error, but by different amount of spatial aliasing. The four upper diagrams show an increased level of the 50 node rendering and thus demonstrate the spatial aliasing impact. The bottom diagrams depict the BEMA treated signals that are mostly slightly improved compared to the non-treated signals. However, some signals, for example the right ear signal for 90°, is even worse than the raw array rendering. In prior informal listening tests, it turned out that BEMA mostly yields unsatisfying results. Especially for highly reverberant rooms, BEMA leads to more audible artifacts than improvements. Nevertheless, it was investigated as part of the listening test.

2.3.5.5 Global Equalization Filter - GEQ

Except for the BEMA algorithm, all approaches either treat the truncation or the spatial aliasing error. To compensate for the spectral artifacts induced by spatial aliasing and SH order truncation, a global undersampling compensation filter (GEQ) was derived. For this, the SHF and the generative first-order low-pass AEQ (Sec. 2.3.4.1) are combined. Figure 2.18 depicts AEQ and SHF (left), as well as the resulting GEQ (right). The high-shelf shape of the GEQ makes clear that the low-pass characteristic of the truncation error has more impact than the high-pass characteristic of spatial aliasing. An important note at this point: The SHFs just depend on the SH rendering order N, but the design of the AEQ on the spatial aliasing frequency (approximated by Eq. 2.21) and thus the sampling scheme order N_{sg}. Hence, two binaural renderings, one up to SH order 3, another up to order 5, requires the same AEQ, but different SHF and thus different shapes of the global equalization filter.

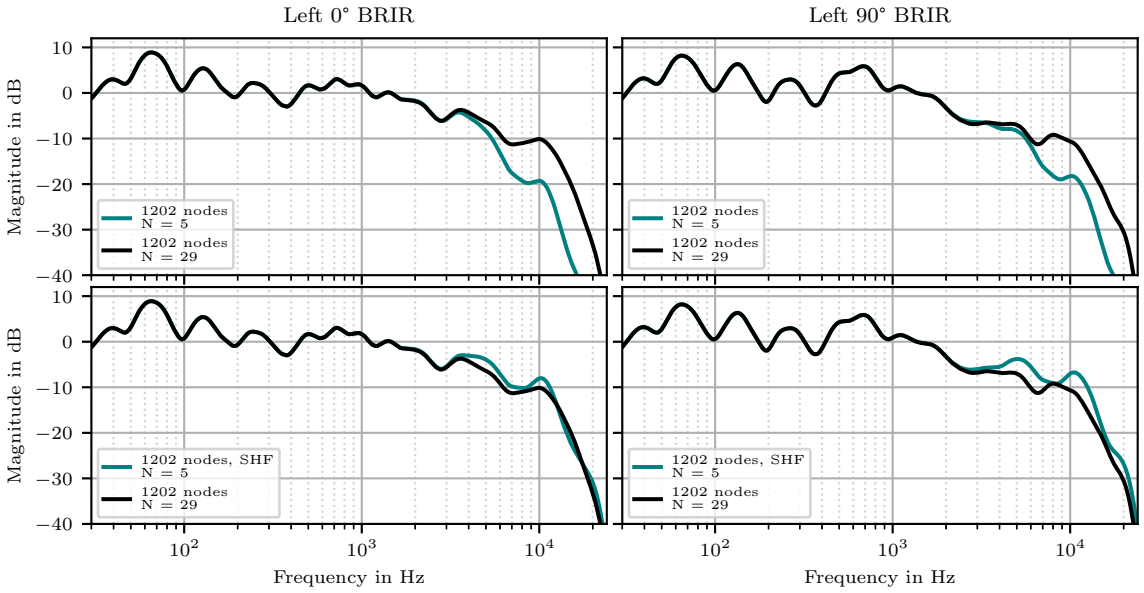

Figure 2.19: SHF: Frequency responses of 0° (left) and 90°(right) ear signals to the left ears resulting from array renderings based on a 1202 sampling node grid rendered up to an SH order N of 29 (black curves) and 5 (blue curves). The upper diagrams show the influence of the truncation error for two different directions. The bottom diagrams depict the improvements of Spherical Head Filter equalization. It can be seen that global equalization achieves different improvements with respect to the direction.

2.3.5.6 Overall Comparison

The previous discussion of the improvement algorithms focuses either on the truncation or spatial aliasing reduction. Therefore, certain rendering configurations were used to illustrate the influence on the single effects. However, in real-life, rendering with an SH order of 5 whereas 1202 microphone signals are given is not meaningful. Moreover, the listening test will investigate the overall perceived improvement of array renderings. To do this, they will be compared to dummy head measurements. Figure 2.23 and Figure 2.24 compare 3rd, 5th and 7th order array renderings based on array measurements in the SBS to corresponding dummy head BRTFs. Figure 2.23 compares BRTFs for 0° orientation to the left ear, Figure 2.24 the 90° BRTFs [3]. The black curve depicts the dummy head BRTF, the dashed grey curve the raw untreated array rendering. Considering the 3rd order rendering at 0°, it can be found that the Tapering and SHF lead to a too high gain of higher frequency components. MagLS and GEQ match the dummy head frequency response the best. For SH order 5 and 7, it can not be clearly determined which algorithm generates the best match of the frequency response. In addition to the frequency responses, the spatial alias frequency according to Equation 2.21 is marked. Consistent with the derivations in Section 2.3.1 this frequency is shifted to higher frequencies for higher-order renderings. Furthermore, it can be seen that the deviation of raw rendering and dummy head reference decreases for higher order renderings.

[3]Similar comparisons for the LBS which was also tested in the listening experiment can be found in the appendix, see. Figure A.3 and Figure A.3.

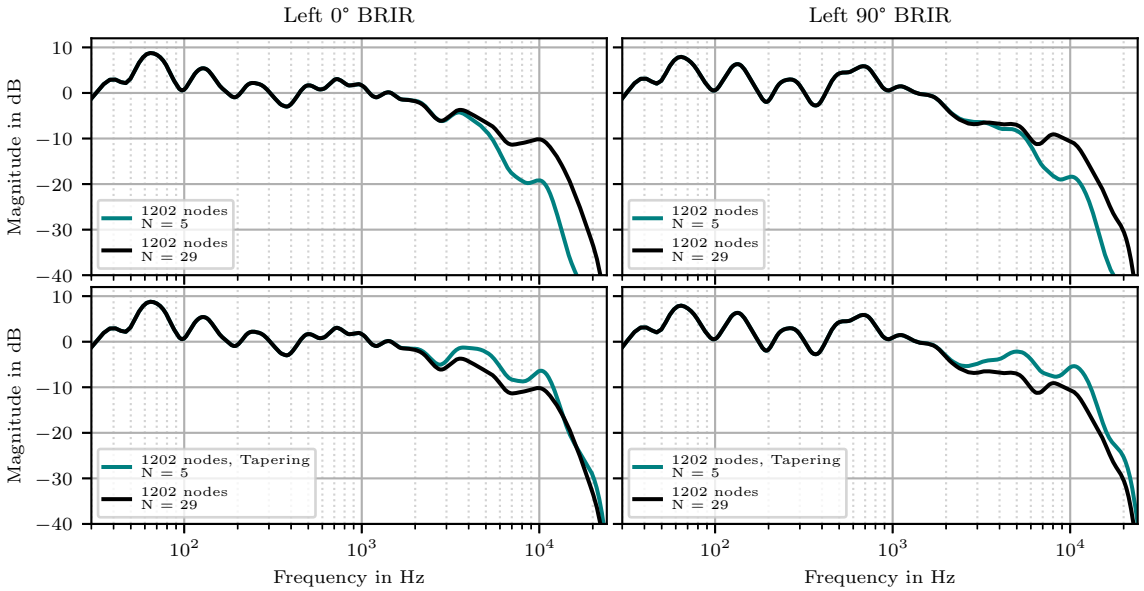

Figure 2.20: Tapering: Frequency responses of 0° (left) and 90° (right) ear signals to the left ear resulting from array renderings based on a 1202 sampling node grid rendered up to an SH order N of 29 (black curves) and 5 (blue curves). The upper diagrams show the influence of the truncation error for two different directions. The bottom diagrams depict the improvements of Tapering equalization. Similar as for the SHF, Tapering achieves different improvements with respect to the sound source direction.

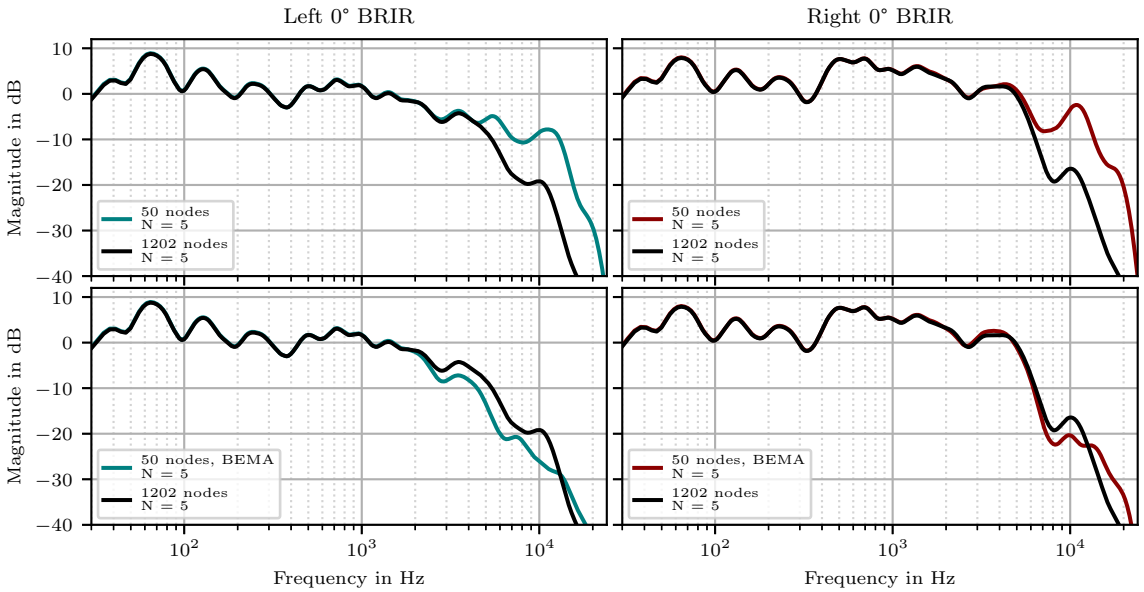

Figure 2.21: BEMA: 0° BRTFs rendered on SH order $N = 5$ based on 1202 (black) and 50 (colored) sampling node grids in the CR7. The upper diagrams illustrate the influence of spatial aliasing, the bottom diagrams the BEMA improvement. It can be seen that BEMA has a different impact on the ipsilateral right (red) and contralateral left (blue) ear signal.

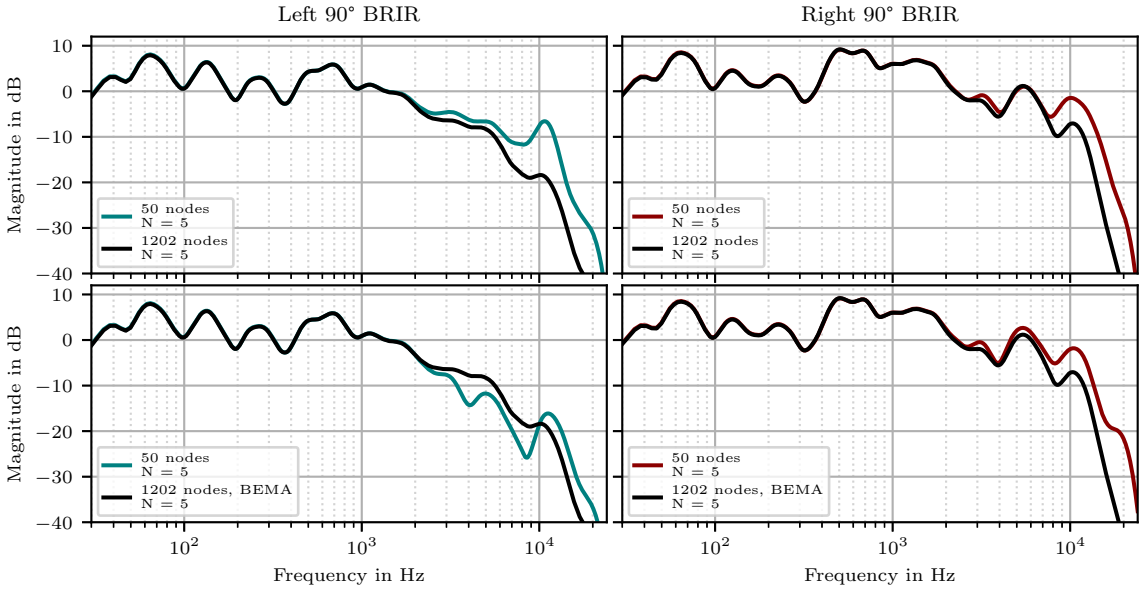

Figure 2.22: BEMA: 90° BRTFs rendered on SH order $N = 5$ based on 1202 (black) and 50 (colored) sampling node grids in the CR7. The upper diagrams illustrate the influence of spatial aliasing, the bottom diagrams the BEMA improvement. Involving Figure 2.21 shows that BEMA achieves different improvements with respect to the direction.

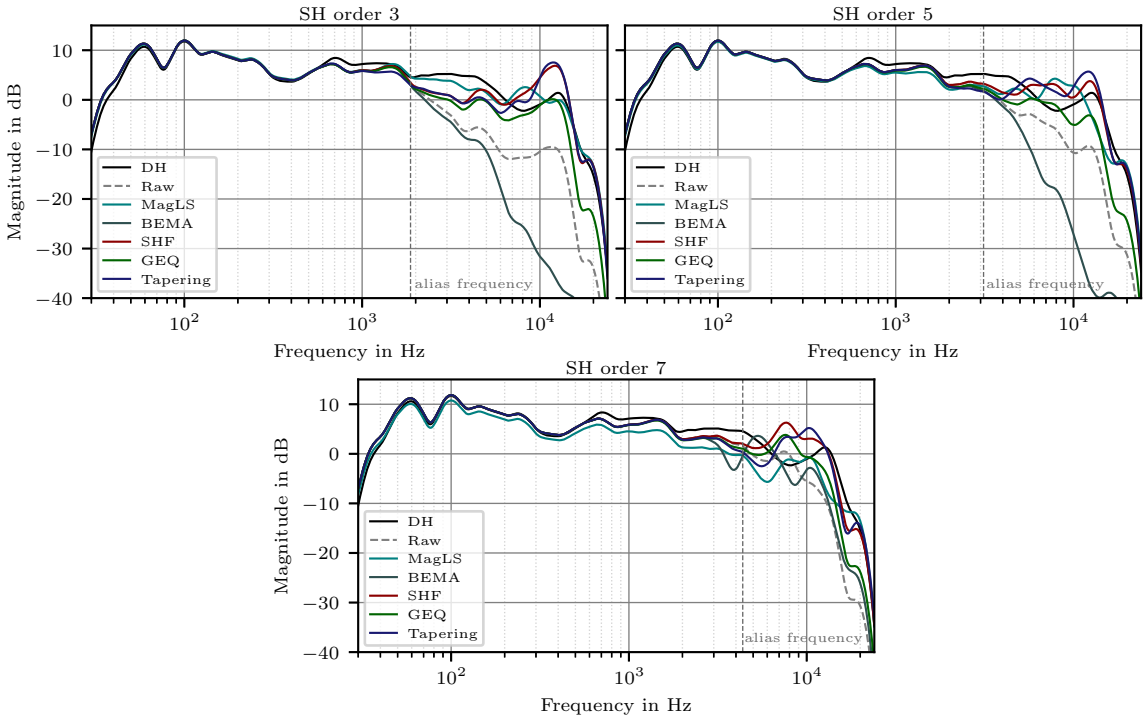

Figure 2.23: Comparsion of 0° BRTFs to the contralateral left ear resulting from 3rd, 5th and 7th order binaural array renderings and dummy head (DH) measurements in the SBS. The array renderings were processed with each of the proposed improvement algorithms.

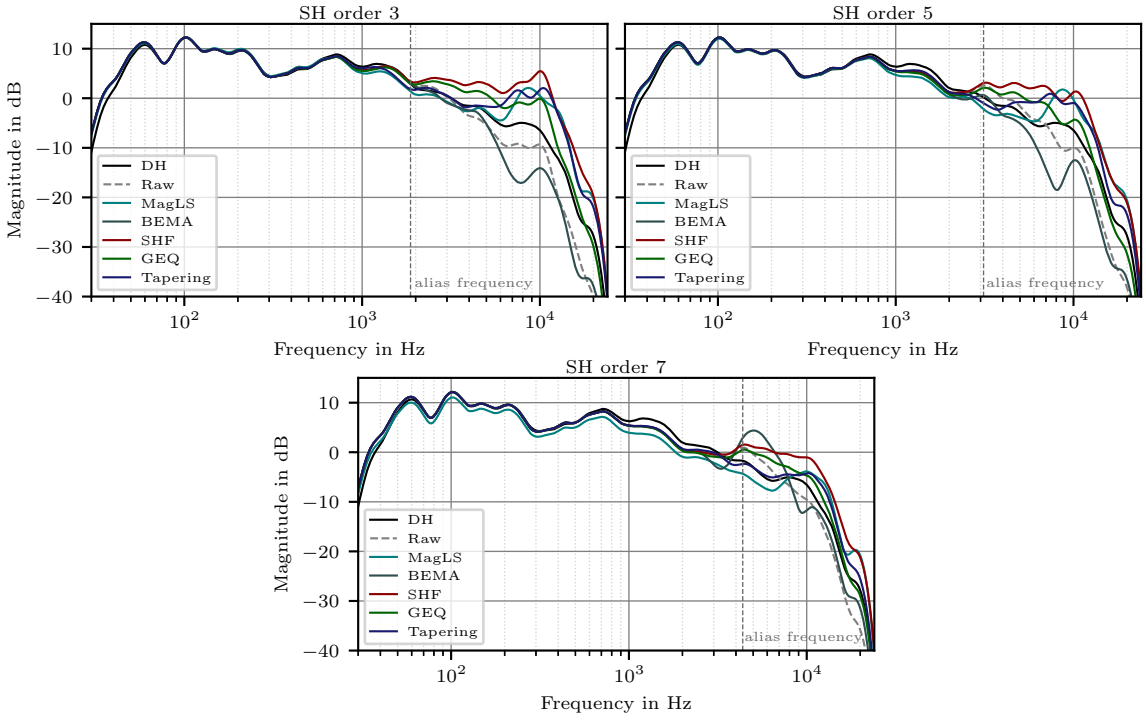

Figure 2.24: Comparsion of 90° BRTFs to the contralateral left ear resulting from 3rd, 5th and 7th order binaural array renderings and dummy head (DH) measurements in the SBS. The array renderings were processed with each of the proposed improvement algorithms.

3
Implementation

3.1 Rigid Sphere Impulse Response Measurements

Although the main purpose of microphone array renderings is dynamic real-time auralization, this work focuses on static, impulse response based renderings. The entire development, as well as the preparation of the listening experiment stimuli, is based on renderings of measured array impulse responses.
Stade et al. (2012) provide a comprehensive measurement database which is used by a large number of researchers studying spherical microphone array processing. Stade et al. (2012) captured dummy head room impulse responses and spherical microphone array impulse responses at the same position in four different rooms at the WDR broadcasting studios in Cologne. The dummy head captures were performed using a Neumann KU100 on a pure horizontal grid with an angular resolution of 1°. The array measurements were done using the VariSphere Array (VSA) developed by Bernschütz et al. (2010). It is a fully automated robotic measurement system which allows to sequentially capture impulse responses for simulating a rigid sphere microphone array. Arbitrary amounts of microphone positions on arbitrary sampling schemes are provided. Stade et al. (2012) captured array responses on a 50, 86, 110, and 1202 sampling point Lebedev grid in a large and small Broadcasting Studio (LBS, SBS), as well as in the Control Room 1 and 7 (CR1, CR7). For all sampling schemes, the microphone capsules were mounted on a rigid sphere body extension with a radius of 8.75 cm. For a more detailed description of the measurement setup, as for example specifications of the microphone capsules, the loudspeaker used for excitation, or the exact measurement positions, the reader is referred to Stade et al. (2012).
Furthermore, Bernschütz (2013) provides an HRTF dataset measured on a 2702 sampling node Lebedev grid which is used for all array renderings in this work. This measurement was done using the KU100 dummy head as well.

3.2 Signal Processing

A complete signal processing chain for binaural auralization of array impulse responses was firstly implemented by Bernschütz et al. (2011). For this, the MATLAB Sound Field Analysis Toolbox (SOFiA) was developed. It allows to render binaural signals for any direction based on array impulse responses measured on arbitrary sampling schemes. Besides a number of additional valuable tools, e.g. for simulating or plotting sound field data, all necessary signal processing steps, introduced in

Section 2.1 were implemented. Hohnerlein and Ahrens (2017) ported the SOFiA toolbox to Python. All signal processing presented in this work, is based on this Python port. A basic overview of the signal flow of the binaural reproduction of array impulse responses is shown in Figure 3.1. The microphone impulse responses measured on a M_{sg}^{array} sampling grid are transformed to the frequency domain first, and then using the DSFT (Eq. 2.19) to the SH domain with an SH order of N. According to the array setup, basically involving the choice of an open or rigid sphere configuration, as well as the array radius, the $N+1$ radial filters are calculated in the frequency domain. Likewise, the M_{sg}^{HRIR} sampling node HRIR dataset is transformed to the SH domain with the same SH order N. The reproduction of the binaural signals is done for certain directions. In case of a dynamic auralization using the SSR, 360 BRTFs on a horizontal grid, denoted as the decomposition grid, are calculated. This calculation is realized using Equation 2.18 by weighting the sound field SH coefficients with the HRTF SH coefficients in the SH domain. A final IFFT yields the binaural impulse responses used for convolution in the SSR.

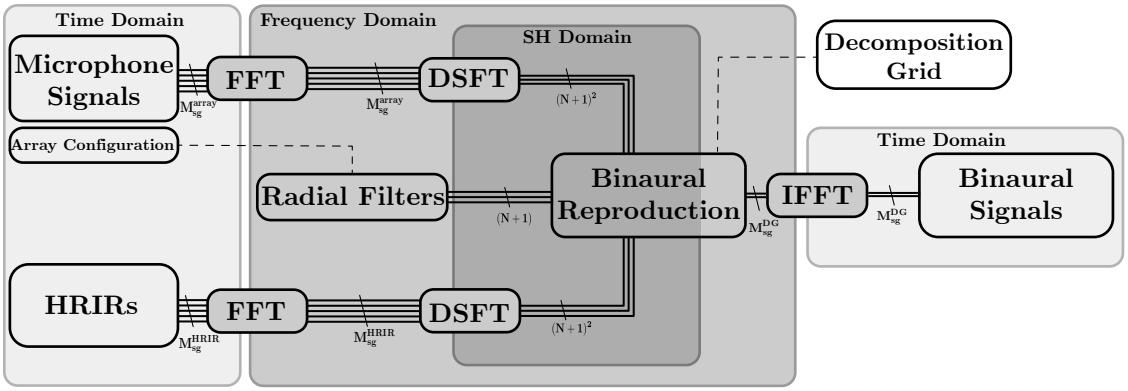

Figure 3.1: Basic signal flow of binaural renderings of microphone array signals. Array impulse responses measured on a M_{sg}^{array} sampling point grid, as well as the M_{sg}^{HRIR} HRIR dataset are transformed to the frequency domain first and then to the SH domain by means of the DSFT up to an SH order N. With respect to the array configuration, $N+1$ radial filters are calculated in the frequency domain. Using Equation 2.18, the binaural signals are calculated for M_{sg}^{DG} directions defined by the decomposition grid. A final IFFT yields the time domain BRIRs.

The existing sound field analysis Python toolbox was extended during the work of this master's thesis. BEMA was directly ported from the SOFiA toolbox. The tapering algorithm was implemented according to Hold et al. (2019)[1]. To realize the MagLS algorithm, the authors (Zaunschirm et al., 2018) provided their MATLAB code. This enabled the possibility to prepare the KU100 HRIR set for renderings on arbitrary SH orders. The Spherical Head Filters, as well as the overall spectral equalization filters, were realized as linear phase FIR filters generated with the firwin function available in the Python-Scipy library. Furthermore, the matrix regularization approach was implemented, however, it turned out that at this stage of development it yields no perceptional satisfying results.

[1] `https://github.com/chris-hld/spaudiopy/tree/ICASSP2019`

3.3 Auralization

The auralization of array data based-BRIRs, as well as the measured dummy head-BRIRs, is done with the SoundScape Renderer (SSR). It allows the real-time convolution of arbitrary HRIRs or BRIRs defined on a pure horizontal grid with 1° resolution with arbitrary input audio signals. The dry audio signal to auralize can be played back on any audio player and routed into the SSR via Jackaudio[2]. To consider the head orientation of the listener, the SSR provides the connection of a variety of head-tracking systems. This enables the presentation of a dynamic binaural synthesis where the entire sound scene keeps static while moving the head. The SSR submits the possibility to load multiple BRIR data sets into the working memory and switch them dynamically during the playback of the dry audio signal. Furthermore, the SSR can handle commands sent via TCP which allows to remotely control the SSR with arbitrary applications. The conducted listening experiment introduced in the next section is based on a Python application with QT5 graphical user interface that controls the SSR via TCP connection.

Figure 3.2: The graphical user interface of the SoundScape Renderer in the Binaural Room Synthesis (BRS) mode shows an example scene consisting of four sound sources. Every sound source represents one BRIR set loaded into the working memory. Every BRIR set can be unmuted dynamically which enables the real-time rendering of a dynamic binaural synthesis.

[2] `http://jackaudio.org/`

4

Perceptual Evaluation

4.1 Introduction

Ideally, spherical microphone array recordings and subsequent binaural reproductions should realize perceptual similar auralizations achieved with dummy head measurements in the same scenario. Bernschütz (2016), or Ahrens and Andersson (2019) conducted listening experiments comparing headphone-based binaural syntheses using dummy head BRIR measurements to headphone-based auralizations based on microphone impulse response measurements with different amounts of microphones. Both, dummy head and array impulse response measurements were held under the same conditions. Bernschütz (2016) found that for SH rendering orders above order 8 most of the perceptual differences decrease significantly. Ahrens and Andersson (2019) confirmed this finding in their follow-up study. Rendering orders up to 8 yield audible differences compared to the dummy head auralizations. In particular, for lateral source directions, remarkable differences arise. The latter can be explained with the truncation error. Further, the investigators show that spatial aliasing has a significant influence on the auralization quality. The present investigation can be seen as a continuation of these studies:

Section 2.3.4 introduced a number of truncation error and spatial aliasing error mitigation approaches that were developed over the last few years. Section 2.3.5 discusses the most promising algorithms, namely MagLS, Tapering, SHF and GEQ, and illustrates their influence on the binaural renderings. This work presents a comparison of the perceivable improvements that can be achieved with this selection of methods by means of a listening experiment. The influence of the spatial aliasing and truncation impairments will not be treated separately, but the overall quality improvement of array renderings compared to corresponding dummy head measurements will be tested.

4.2 Experimental Design

4.2.1 Stimuli

To compare the algorithms in a listening experiment, different BRIR renderings were presented to the listener using a dynamic binaural synthesis. Each BRIR was rendered based on array impulse responses in the WDR broadcasting studios, as

described in Section 3. Ahrens and Andersson (2019) showed that up to an SH rendering order of 8 audible differences arise. As this study aims at investigating undersampled array constellations, renderings of orders 3, 5 and 7 are compared. The rendering orders 3 and 5 are based on the 50 sampling node Lebedev grid measurement, the 7th order rendering on the 86 sampling node Lebedev grid. To study the dependency on the environment, the binaural renderings were done for array impulse responses in the LBS (approx. $RT_{60} = 1.8$ sec, at 1 kHz) and SBS (approx. $RT_{60} = 1.1$ sec, at 1 kHz). Each of these rendering configurations were processed with each improvement algorithm. It was presented a dynamic binaural synthesis where the participant was free to rotate its head and listen to all directions. To simulate an initial lateral head orientation which allows for studying the direction-dependent performance of the algorithms, each BRIR was rotated clockwise for 90°. This is of high interest in particular for the algorithms reducing the truncation error. To ensure consistency to the study of Ahrens and Andersson (2019), for all renderings the radial filters were soft-limited to 0 dB[1].
The use of linear-phase filters results in BRIRs which are time delayed in relation to the BRIRs measured with the dummy head. To provide a continuous playback while switching the stimuli during the listening experiments, all delays where compensated. This was achieved by shifting the BRIRs according to their peak measured with a simple onset detection provided by the AKtools MATLAB toolbox[2] (Brinkmann and Weinzierl, 2017).
As the test signal, acoustical dry drum recordings were presented. Drum recordings consist of a wide spectrum and typically provide high transients that make drums to rather critical test signals. Furthermore, Ahrens and Andersson (2019) used the same test signal which supports the consistency of the experimental results.

4.2.2 Experimental Paradigm

To compare the different algorithms, a listening test based on the multiple stimulus with hidden reference and anchor (MUSHRA) test paradigm was used. The MUSHRA methodology is a multi-stimulus test design proposed by the International Telecommunication Union (ITU), see ITU-R BS.1534-3 (2015). It was developed for evaluating the preceptual quality of audio systems and codecs. It allows to compare up to 12 different systems, or different levels of audio systems presented at the same time. Therefore, the reference stimulus labeled as such, as well as multiple further stimuli were presented per trial. One of the stimuli is the hidden reference, one of them a hidden anchor-stimulus. All stimuli are compared to the reference stimulus in terms of intermediate perceived quality. The anchor is a stimulus that should be perceived as worst by the listener and is used for screening of the ratings. Each rating is done with a continuous movable slider ranging from 0-100 which allows to register very small differences. The sliders are divided into five evenly spaced segments labeled with 'Excellent' (100-80), 'Good' (80-60), 'Fair' (60-40), 'Poor'

[1] Again the reader is referred to the appendix where the soft-limited radial filters are depicted, Figure A.1

[2] `www.ak.tu-berlin.de/aktools`

(40-20), 'Bad' (20-0).
In contrast to the MUSHRA design introduced by the ITU, the aim was to evaluate the overall perceived difference to the reference stimuli instead of their intermediate quality. Therefore, a slightly adapted test design was used in which the slider labels were replaced with 'No difference' (100), 'Small difference' (75), ' Moderate difference' (50), 'Significant difference' (25) and 'Huge difference' (0). Furthermore, the placements of the labels were adjusted such that 'No' and 'Huge' difference represents the extreme ratings 100 and 0. Figure 4.1 depicts the PyQT based experimental graphical user interface developed for the listening test.
All stimuli described above have to be compared to the corresponding dummy head auralization in terms of overall perceived differences. For the anchor stimulus, a 3 kHz low-pass filtered diotic signal based on the associated dummy head BRIR was generated. Hence, for each trial, BRIRs processed with the algorithms MagLS, BEMA, Tapering, SHF, GEQ, as well as the untreated BRIR rendering (raw), the hidden dummy head reference, and the anchor stimulus were presented. These conditions were rendered for two rooms, three SH orders, and two directions. Table 4.1 summarizes these 12 trials with 8 conditions each.

Table 4.1: Overview of the stimuli tested in the final listening experiment. All of the conditions were presented per trial, whereas the reference stimulus (DH Ref), anchor stimulus, and the raw rendering are listed as an algorithm. Each of those sets of conditions were rendered for 3 SH orders, 2 directions and 2 rooms. In total 12 trials with 8 conditions each.

Algorithms	SH orders (grid)	Directions	Rooms
BEMA	3 (50)	0°	LBS
MagLS	5 (50)	90°	SBS
SHF	7 (86)		
Tapering			
GEQ			
Raw			
DH Ref			
Anchor			

4.2.3 Setup

The experiment was conducted in a quiet acoustically damped audio laboratory at the Chalmers University. The participants were presented a dynamic binaural synthesis using AKG K702 headphones with a Lake People G109 headphone amplifier at a playback level of 66 dB(A). The head orientation was tracked with a Polhemus Patriot. The rendering of the binaural VAE was done with the SSR in BRS mode. A Python application with QT5 graphical user interface which controls the SSR via TCP commands was developed for the listening experiment. The output signals of the SSR were routed to an Antilope Audio Orion 32 channel DA converter via a 48 kHz JACK-audio server providing 512 samples per buffer. To compensate for the binaural chain of the AKG headphones and the KU100 dummy head, a headphone-equalization according to Stade et al. (2012) was performed. The entire rendering and performance of the listening experiment was done on an iMac Pro 1.1.

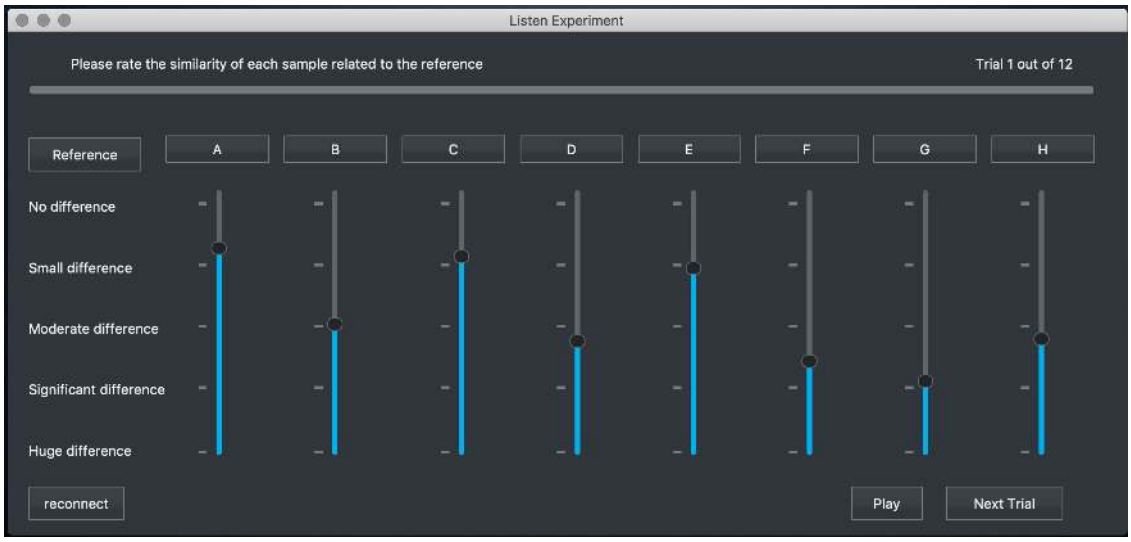

Figure 4.1: PyQT5 implementation of the graphical user interface for the listening test. 8 stimuli, one of them the hidden reference, one the hidden anchor, are compared to the reference stimulus. Each slider allows to rate the perceived difference to the dummy head reference.

4.2.4 Procedure

12 participants, 3 of them female, aged between 22 and 50 took part in the experiment. Most of them were master's students or staff at the Division of Applied Acoustics of the Chalmers University. 9 participants had experience with binaural synthesis and 8 reported already participated in a listening experiment before. The subjects sat in front of a computer screen with a keyboard and mouse. Their task was to rate the similarity of each stimulus compared to the reference according to the stimuli preparation discussed above. Each stimulus, as well as the reference could be listen to as much as desired. The participants were allowed and strongly encouraged to move their heads during the presentation of the stimuli. At the beginning of each experiment, the subjects had to rate four training stimuli which cover the entire similarity range of the presented stimuli in the following test. It took about 30 minutes for the participants to rate all 12 trials.

4.3 Results

According to the recommendation of ITU-R BS.1534-3 (2015), all anchor and reference ratings were post-screened before applying statistical analysis. All anchor ratings were below 30 and most of the reference ratings higher than 80. Solely two

reference ratings of 50 and 49 were conspicuous which, however, is a relatively low portion of in total 96 ratings per participant. There were no further noticeable abnormalities in these ratings, wherefore none of the participants was excluded from the statistics.

A first overview of the results is presented as boxplots in Figure 4.2 illustrating the ratings for the SBS and LBS at 0° and 90° separately. Each of the four figures shows a boxplot for each algorithm and the SH orders 3 (red), 5 (black), and 7 (blue). The boxes displaying the 25th and 75th percentiles and the corresponding median value as a black line. Furthermore, outliers are indicated as grey circles and minimum and maxim ratings, not identified as outliers, by the upper and lower whiskers. Boxplots refer to median values which are presented in Table A.1 in the appendix. Additionally, the table shows a number of further descriptive statistical values.
The boxplots confirm that for most conditions the anchor and reference stimuli were indicated as such by the listener, the anchor median value of all conditions is 1 and the reference rating median value is 100. Furthermore, the boxplots for the untreated renderings illustrate that for all four room-direction combinations, higher order renderings yielded higher similarity ratings. The 5th order rendering consequently achieved higher ratings than the 3rd order rendering, however, 7th order renderings in the SBS achieved lower ratings than the rendering on SH order 5. Apart from this, it is remarkable that all improvement algorithms achieved higher ratings than raw renderings except the BEMA algorithm. Hence, all other algorithms achieved perceptual improvements.

For a more detailed analysis, repeated measures ANOVAs were performed. To test the data requirements for the ANOVA a Lilliefors test for normality distribution was applied. It failed to reject the null hypothesis in 5 of 72 conditions at a significance level of .05. However, parametric tests as ANOVAs are generally robust to violations of normal distribution assumption. For the further analysis, Greenhouse-Geisser corrected p-values are considered (the associated ϵ values for correction of the degrees of freedom of the F-distribution are reported as well).
A four-way repeated measures ANOVA with the within-subject factors algorithm (BEMA, MagLS, Tapering, SHF, GEQ and raw), order N (3, 5, 7), room (SBS, LBS) and direction (0° , 90°) was performed. Even though the denotation is rather inappropriate, the raw rendering is one factor-step of the within-subject factor algorithm. The ANOVA results are presented in Table 4.2.
In addition, a number of nested repeated measures ANOVAs were performed whose results are presented in the appendix. For each of the six algorithms, one three-way ANOVA with the factors order (3, 5, 7), room (SBS, LBS), and direction (0° , 90°), as well as a four-way ANOVA with the subset of MagLS, Tapering, SHF, and GEQ for the factor algorithm and the factors order (3, 5, 7), room (SBS, LBS), and direction (0° , 90°) was applied.

Since ANOVAs refer to mean values, Figure 4.3 was generated to support the ANOVA results. It depicts meanplots for each algorithm and the orders 3 (red), 5 (black), and 7 (blue), separately. Each mean value was calculated by averaging

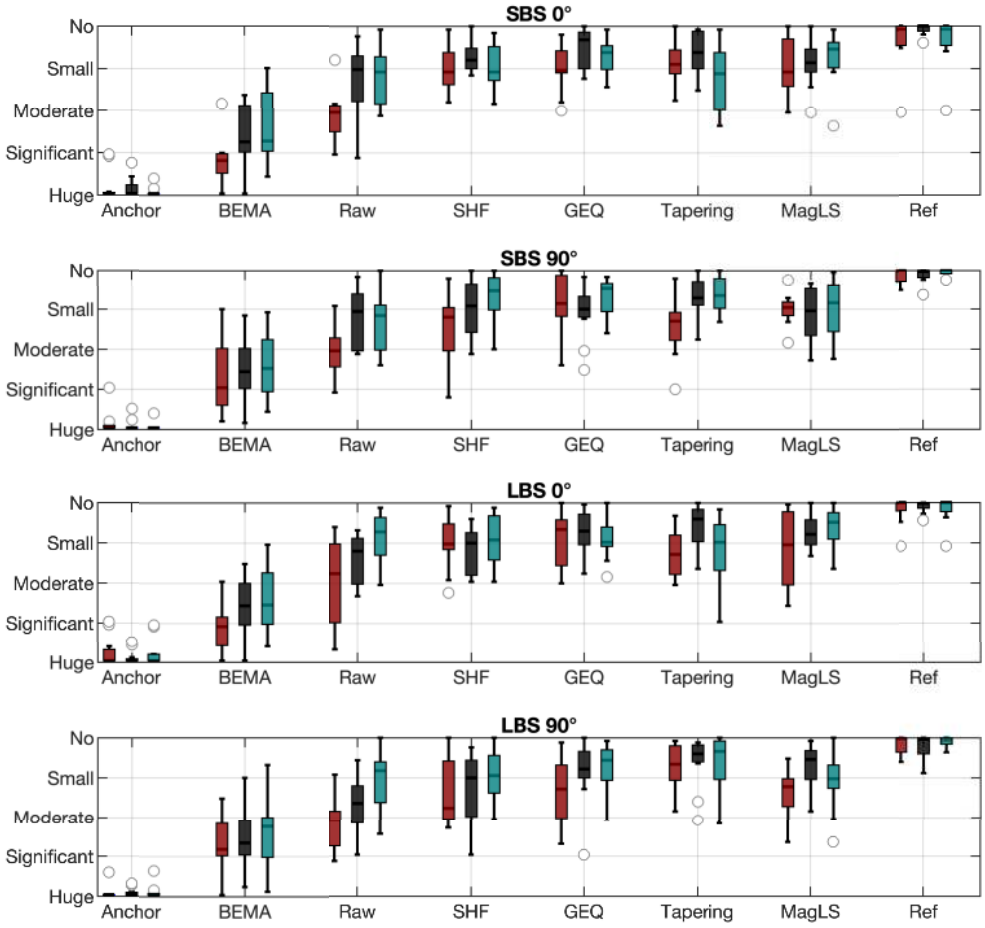

Figure 4.2: Boxplots illustrating the ratings for each room and direction separately. Each figure shows the boxplots for each algorithm and the SH orders 3 (red), 5 (black) and 7 (blue). Each box indicates the 25th and 75th percentiles and the median value as a black line. The outliers are marked as grey circles and the minimum and maximum ratings not identified as outliers with black upper whiskers.

the ratings for both rooms, both directions, and all participants. Furthermore, 95% within-subject confidence intervals as proposed by Loftus (2002) and Jarmasz and Hollands (2009) based on the algorithm main effect are shown. The figure confirms the observations taken from the boxplots that for the raw rendering higher SH orders achieved better ratings than lower-order renderings and that, except for BEMA all algorithms led to an improvement compared to the raw auralization. More clearly than the boxplots, the meanplots suggest that the perceptual quality of the array rendering does not scale linearly with the SH order. Considering just the raw condition shows that the mean ratings for 5th and 7th order renderings are more closer to each other than the 5th and 3rd order ratings.
Similar to the boxplots, the meanplots indicate that the 7th order renderings involving the Tapering algorithm were rated worse compared to the 5th order Tapering renderings. Additionally, it can be seen that the average ratings of the algorithms SHF, GEQ, Tapering and MagLS are located in the same range. That leads to the

Table 4.2: Results of the four-way repeated measures ANOVA with the within-subject factors algorithm (BEMA, MagLS, Tapering, SHF, GEQ, Raw), order (3, 5, 7), direction (0°, 90°), and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Algorithm	5	97.964	.572	.899	$< .001^*$	$< .001^*$
Order	2	27.681	.808	.716	$< .001^*$	$< .001^*$
Direction	1	1.014	1	.084	.335	.335
Room	1	.462	1	.040	.511	.511
Algorithm×Order	10	4.989	.514	.312	$< .001^*$	$< .001^*$
Algorithm×Direction	5	3.302	.703	.231	$.011^*$	$.024^*$
Order×Direction	2	1.811	.876	.141	.187	.193
Algorithm×Room	5	1.267	.683	.103	.291	.300
Order×Room	2	.224	.677	.020	.801	.715
Direction×Room	1	.016	1	.001	.902	.902
Algorithm×Order×Direction	10	2.329	.518	.175	.016	.052
Algorithm×Order×Room	10	1.817	.481	.142	.066	.128
Algorithm×Direction×Room	5	1.544	.575	.123	.191	.223
Order×Direction×Room	2	.372	.917	.033	.693	.676
Algorithm×Order×Direction×Room	10	1.589	.424	.126	.119	.190

ϵ_{GG}: Greenhouse-Geisser epsilons
p_{GG}: Greenhouse-Geisser corrected p-values.
Statistical significance at 5% level are indicated by asterisks

assumption that these four algorithms achieve similar auralization improvements. Their average ratings for all SH rendering orders are located between 69.65 and 84.46. (GEQ: 74.44 ($N = 3$), 80.52 ($N = 5$), 81.02 ($N = 7$), SHF: 69.65 ($N = 3$), 73.48 ($N = 5$), 78.27 ($N = 7$), MagLS: 72.4 ($N = 3$), 78.67 ($N = 5$), 79.29 ($N = 7$), Taper: 73.75 ($N = 3$), 84.46 ($N = 5$), 77.73 ($N = 7$).

The following analysis refers to main effects and first order interactions only, and neglects the three and four-way interaction effects. For the four-way ANOVA involving all algorithms it was found that the algorithm and order main effects, as well as the first order interaction effects algorithm × order and algorithm × direction were significant. These significances will be examined successively in the following:

Algorithm and Order Dependency

The significances of the main effects order ($F(2, 22) = 27.681$, $p < .001$, $\eta_p^2 = .716$, $\epsilon = .808$) and algorithm ($F(5, 55) = 97.964$, $p < .001$, $\eta_p^2 = .899$, $\epsilon = .572$) match the findings observed in Figure 4.2 and Figure 4.3. Higher SH orders lead to different, and mostly higher similarity ratings. Hence, the auralization quality highly depends on the rendering order and the algorithm.
The first order interaction effect of the factor algorithm × order ($F(10, 110) = 4.98$, $p < .001$, $\eta_p^2 = .312$, $\epsilon = .514$) shows that not solely the SH order influences the quality of the untreated rendering, but also the algorithm performs differently with respect to the SH order. The single algorithm ANOVAs support this. For every ANOVA the order was indicated as a significant effect.

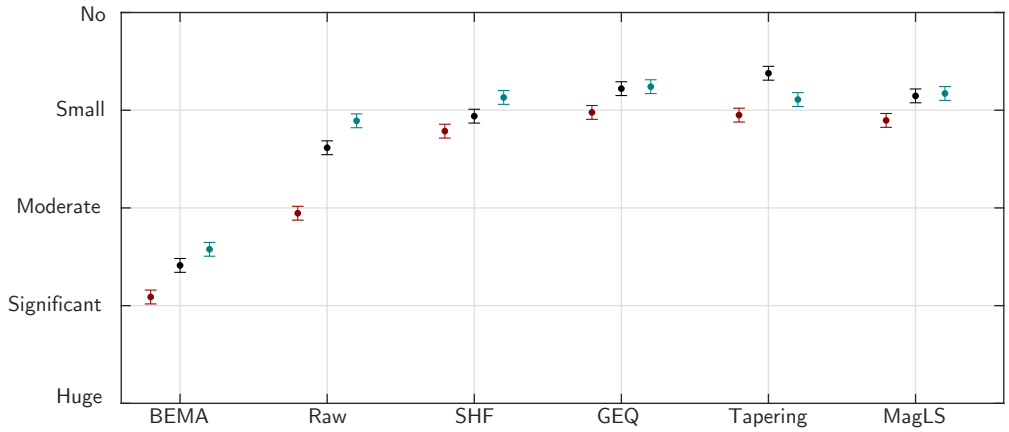

Figure 4.3: Mean values for algorithm and SH order separately. Every mean value was calculated by averaging all room and direction ratings for each participant. The 95% within-subject confidence intervals for the main effect algorithm were calculated according to Loftus (2002) and Jarmasz and Hollands (2009). Red colors indicate order 3 renderings, black colors order 5 and blue order 7 renderings. The highest mean value for order 3 and 7 were achieved by the Global Equalization Filter (74.44, 81.02), for order 5 by the Tapering approach (84.46).

To validate the observation that the algorithms SHF, GEQ, Tapering and MagLS achieved similar improvements, the four-way repeated measures ANOVA just for these algorithms was applied. The factor algorithm was not indicated as significant anymore ($p = .124$). Thus, in general, all algorithms except BEMA achieved similar perceptual improvements.

Directional Dependency

The four-way main ANOVA yielded no significance for the factor direction as main effect ($p = .335$), but it was found that the interaction of algorithm $\times$ direction ($F(5, 55) = 3.6308$, $p < .024$, $\eta_p^2 = .231$, $\epsilon = .703$) was significant. Figure 4.4 was generated to study the directional dependency. It shows separated meanplots divided with respect to the directions 0° (left) and 90° (right). Considering the mean values for the raw rendering shows that 90° conditions consequently achieved slight worse results than the frontal ($N = 3$: 49,75 (0°), 47.54 (90°), $N = 5$: 67.085 (0°), 63.705 (90°), $N = 7$: 73.83 (0°), 70.75 (90°)). The distribution of the order dependent mean values for SHF, Tapering, MagLS, and GEQ behave completely different for both directions. In particular for the 7th order rendering. Correspondingly, the single algorithm ANOVAs for MagLS and Tapering showed a significance for the main effect direction ($F(1, 11) = 6.806$, $p < .024$, $\eta_p^2 = .382$, $\epsilon = 1$), ($F(1, 11) = 5.012$, $p < .047$, $\eta_p^2 = .313$, $\epsilon = 1$) and the four-way ANOVA for the algorithms SHF, GEQ, Tapering, and MagLS a significant interaction of order $\times$ direction ($F(2, 22) = 5.054$, $p < .019$, $\eta_p^2 = .315$, $\epsilon = .911$). This interaction suggests that the algorithms SHF, GEQ, Tapering and MagLS, performed different for lateral or frontal sound sources. In addition, the figure illustrates another interesting finding. For the algorithms SHF, GEQ, and Tapering the 7th order rendering achieved higher ratings at 90° than

the presentation in the front.

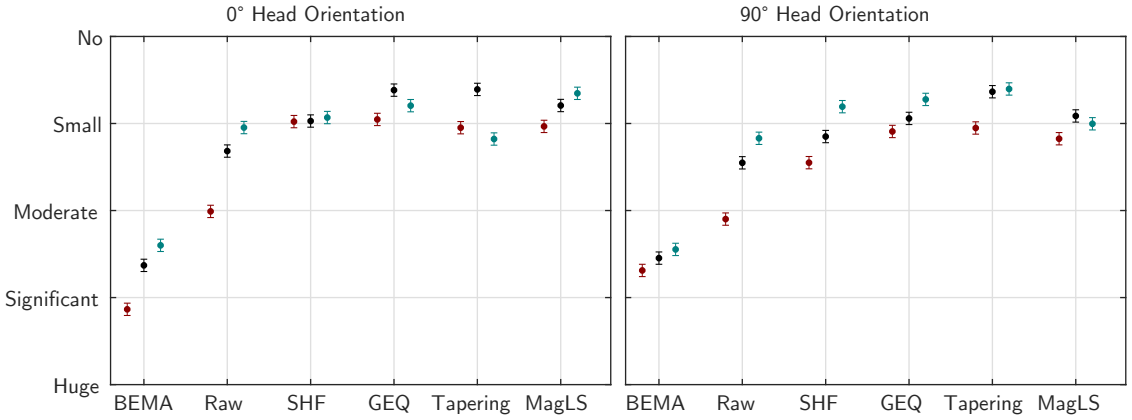

Figure 4.4: Mean values for algorithm and SH order separately divided to 0° (left) and 90° (right) presentations. Every mean value was calculated by averaging all room ratings of all participants. The 95% within-subject confidence intervals were calculated according to Loftus (2002) and Jarmasz and Hollands (2009). Red colors indicate order 3, black colors order 5 and blue order 7 renderings. It remarkable that the algorithms SHF, GEQ and Tapering achieved better results for the 90° condition than for the frontal presentation.

Room Dependency

Neither any of the ANOVAs presented the room as significant main effect, nor any of the 4-way ANOVAs involving multiple algorithms showed a significant interaction effect for the room. Solely the first order interaction order × room ($F(2, 22) = 5.751$, $p < .022$, $\eta_p^2 = .343$, $\epsilon = .678$) for the raw rendering ANOVA, and direction × room ($F(1, 11) = 5.963$, $p < .033$, $\eta_p^2 = .352$, $\epsilon = 1$) for Tapering were indicated as significant. It can thus be concluded that the presented room had no major influence on the rendering quality, or the perceptional improvement of the algorithms in general. Figure 4.5 displays the mean values for each algorithm and SH order for the SBS (left) and LBS (right) separately. It can clearly be seen that for the raw condition the SBS 7th order renderings were rated worse than 5th order renderings which can not be observed for the LBS. For the GEQ or MagLS algorithms it acts exactly the other way round. BEMA and Tapering achieved similar results in both rooms.
Since there is no regularity in the mean value distributions, the plots support the ANOVA results that the algorithms are not significantly influenced by the room.

4.4 Discussion

The results of the presented listening experiment confirm that binaural array renderings that were not further improved with respect to the undersampling error, achieved more perceptual similarity to the dummy head auralization when rendered on higher SH orders. However, even untreated 7th order renderings were rated

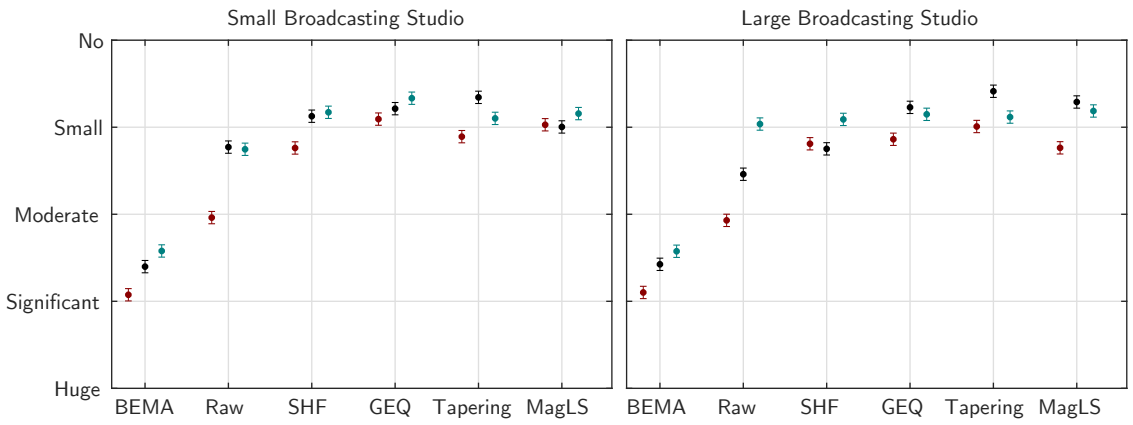

Figure 4.5: Mean values for algorithm and SH order separately divided to SBS (left) and LBS (right). Every mean value was calculated by averaging the ratings for all directions, and all participants. The 95% within-subject confidence intervals were calculated according to Loftus (2002) and Jarmasz and Hollands (2009). Red colors indicate order 3 renderings, black colors order 5 and blue order 7 renderings.

remarkably different by the listener. This matches the findings of Ahrens and Andersson (2019), or Bernschütz (2016) who showed that array renderings below SH order 8 lead to perceivable differences compared to the dummy head rendering.
A further conclusion is that all algorithms except BEMA achieved an improvement in contrast to untreated renderings. In fact, BEMA led to noticeable impairments, and is therefore excluded from the following discussion. Even Bernschütz (2016, pp. 233-234) observed that BEMA introduces audible artifacts.
In total, the results of the algorithms MagLS, SHF, Tapering and GEQ are located in the same range, hence they all led to similar level of improvements. Further, it was found that all algorithms performed different for frontal and lateral sound source directions with respect to the rendering order. Thus, the results do not confirm that approaches based on a global equalization, as the SHF and GEQ, are more vulnerable to lateral impairments. The room was not identified as an important factor. There are some differences in the ratings for the LBS and SBS, however, no regularity can be exposed. The same was found by Ahrens and Andersson (2019) and Bernschütz (2016). They also observed that the room has no significant influence on the quality of binaural array renderings.

The conspicious SH order dependent effect that 7th order renderings for lateral sound sources were rated better compared to frontal sources for most of the algorithms, is consistent to the findings of Ahrens and Andersson (2019). One possible explanation could be that according to Blauert (1997), the human auditory system is able to locate sound sources in the front with a higher spatial resolution than lateral sources. For that reason, differences in frontal directions might be perceived as more distinctive than differences at lateral directions. This effect becomes more significant with higher auralization quality what could explain the lower ratings for 7th order renderings compared to 5th order renderings, especially for the algorithms Tapering, GEQ or MagLS.

5
Conclusion

5.1 Summary

This work explains and extensively discusses the impairments that arise in binaural renderings of spatial undersampled spherical microphone array data. The perceptual influences of the truncation error and the spatial aliasing effect are revealed, and the state-of-the-art approaches to mitigate these effects are described. The most promising algorithms were implemented and perceptual evaluated in a conclusive listening experiment.
It has been shown that there are some algorithms that significantly improve the rendering of undersampled microphone array renderings. The Magnitude Least-Squares algorithm by Schörkhuber et al. (2018), the Spherical Head Filters by Ben-Hur et al. (2017), the Tapering algorithm by Hold et al. (2019), as well as a Global Equalization Filter based on findings of Bernschütz (2016) all achieved perceptual improvements in contrast to untreated array renderings. Although the listening experiment showed slight differences for varying circumstances as the spherical harmonics rendering order, sound source direction, or the presented room, all these algorithms result in rather similar level of improvements and no algorithm was indicated as most applicable. Even improved 7th order renderings was perceived as different compared to the auralization of dummy head data auralizations.
MagLS and Tapering are appropriate algorithms to reduce the truncation error. For the mitigation of spatial aliasing artifacts no satisfying algorithms were found. BEMA is the only approach that handles the spatial loss introduced by spatial aliasing, instead of just compensating the coloration impairment. However, simulations, as well as the listening experiments show that it leads to more artifacts than improvements. This is caused by the fact that this approach is based on an average value of the captured sound field spherical harmonics coefficients in certain frequency bands to gain spatial information for the aliased components. This work focuses on static, room impulse response based renderings. Logically, the averaging of the spatial energy of an entire room impulse response leads to a spatial concentration at one point that leads to less spaciousness in the binaural reproduction. As already proposed by Bernschütz (2016), a block-based implementation of BEMA could prevent this effect. Instead of concentrating the total sound field energy at one point, processing smaller time signals sequentially could maintain the spaciousness while reducing the spatial aliasing.
Moreover, it was shown that an appropriate equalization as the Spherical Head Filters or the global equalization of the binaural signals yields noteworthy improve-

ments. This lead to the assumption that for the average listener an adequate timbre correction of the presented virtual acoustic environment is completely sufficient. However, from the scientific point of view the presence of spatial aliasing and truncation error and the associated loss of spatial information is intolerable.

5.2 Follow-Up Studies

As briefly mentioned, the conducted listening experiment was done for two quite similar rooms with respect to the reverberation time. Although no strong significant influence of the room was found this has to be evaluated for a higher variety of environmental conditions.
The matrix regularization algorithm introduced by Alon and Rafaely (2012) was implemented, but it did not yield perceptual improvements. However, it offers a promising approach that could be further investigated and enhanced in the future.
The MagLS was indicated as sufficient algorithm for preventing the influence of SH order truncation. However, binaural signals processed with MagLS pre-modified HRTFs are still effected by the high-pass effect of spatial aliasing, as shown in Figure 2.23 and Figure 2.24. Adequate equalization of the binaural signals rendered with MagLS HRTFs could further improve the perceptual quality of binaural array reproductions. In general, the combination of different approaches has neither been investigated nor perceptual evaluated up to now. This might open up an interesting future field of research.

Bibliography

Ahrens, Jens (2012): *Analytic Methods of Sound Field Synthesis*.

Ahrens, Jens and Carl Andersson (2019): "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre." In: , **145**(April), pp. 2783–2794.

Alon, David Lou and Boaz Rafaely (2012): "Spherical microphone array with optimal aliasing cancellation." In: *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2012*, pp. 1–5.

Alon, David Lou and Boaz Rafaely (2017): "Spatial decomposition by spherical array processing." In: *Parametric Time-Frequency Domain Spatial Audio*, pp. 25–47.

Andersson, Carl (2017): "Headphone Auralization of Acoustic Spaces Recorded with Spherical Microphone Arrays." In: .

Ben-Hur, Zamir; David Lou Alon; Boaz Rafaely; and Ravish Mehra (2019): "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions." In: *EURASIP Journal on Audio, Speech, and Music Processing*, **2019**(1).

Ben-Hur, Zamir; Fabian Brinkmann; Jonathan Sheaffer; Stefan Weinzierl; and Boaz Rafaely (2017): "Spectral equalization in binaural signals represented by order-truncated spherical harmonics." In: *The Journal of the Acoustical Society of America*, **141**(6), pp. 4087–4096.

Ben-Hur, Zamir; Jonathan Sheaffer; and Boaz Rafaely (2018): "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction." In: *Applied Acoustics*, **134**(February), pp. 138–144.

Bernschütz, Benjamin (2012): "Bandwidth Extension for Microphone Arrays." In: *AES 133th Convention*, pp. 1–10.

Bernschütz, Benjamin (2013): "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100." In: *Fortschritte der Akustik – AIA-DAGA 2013*, pp. 592—-595.

Bernschütz, Benjamin (2016): "Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording." In: , p. 264.

Bernschütz, Benjamin; Christoph Pörschmann; Sascha Spors; and Stefan Weinzierl (2011): "SOFiA Sound Field Analysis Toolbox." In: *Proceedings of the International Conference on Spatial Audio - ICSA 2011*, pp. 8–16.

Bernschütz, Benjamin; et al. (2010): "Entwurf und Aufbau eines variablen sph ¨ arischen Mikrofonarrays f u ¨ r Forschungsanwendungen in Raumakustik und Virtual Audio Einleitung Dimensionierung und Genauigkeit R ¨ aumliche Aliasartefakte Einbr u ¨ che der Modalen Amplituden Zusammenfassung Lit." In: , pp. 717–718.

Blauert, Jens (1997): *Spatial Hearing*. Camebridge: Hirzel Verlag Stuttgart.

Brinkmann, Fabian and Stefan Weinzierl (2017): "AKtools - an open software toolbox for signal acquisition, processing, and inspection in acoustics." In: *142nd Convention Audio Engineering Society*, (i), pp. 1–6.

Brinkmann, Fabian and Stefan Weinzierl (2018): "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition." In: *Conference on Audio for Virtual and Augmented Reality*, pp. 1–10.

Dziwis, D.; Tim Lübeck; Johannes M. Arend; and Christoph Pörschmann (2019): "Development of a 7th Order Spherical Microphone Array for Spatial Audio Recording Development and Implementation Sphere." In: *Fortschritte der Akustik – DAGA 2019*, (April), pp. 883–885.

Elahi, Usama; Zubair Khalid; and Rodney A Kennedy (2019): "Spatially Constrained Anti-Aliasing Filter Using Slepian Eigenfunction Window on the Sphere." In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Geier, Matthias; Jens Ahrens; and Sascha Spors (2008): "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods." In: *Proceedings of 124th Audio Engineering Society Convention 2008*, pp. 179–184.

Gonzalez, Raimundo; Joshua Pearce; and Tapio Lokki (2018): "Modular Design for Spherical Microphone Arrays." In: *Proceedings of the AES Conference on Audio for Virtual and Augmented Reality*, (August).

Helmholz, Hannes; Carl Andersson; and Jens Ahrens (2019): "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals." In: *DAGA 2019: 45. Deutsche Jahrestagung für Akustik*, pp. 2–5.

Hohnerlein, Christoph and Jens Ahrens (2017): "Spherical Microphone Array Processing in Python with the sound field analysis-py Toolbox." In: *Fortschritte der Akustik – DAGA 2017*, pp. 1033–1036.

Hold, Christoph; Hannes Gamper; Ville Pulkki; Nikunj Raghuvanshi; and Ivan J. Tashev (2019): "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation." In: , **2**(3), pp. 261–265.

ITU-R BS.1534-3 (2015): "Method for the subjective assessment of intermediate

quality level of audio systems." In: *International Telecommunication Union*, **3**, p. 34.

Jackson, John David (1962): *Classical Electrodynamics*. Illinois: John Wiley & Sons, Inc.

Jarmasz, Jerzy and Justin G. Hollands (2009): "Confidence Intervals in Repeated-Measures Designs: The Number of Observations Principle." In: *Canadian Journal of Experimental Psychology*, **63**(2), pp. 124–138.

Kinsler, Lawrence.E; Austin.R Frey; Alan.B Coppens; and James.V. Sanders (1999): "Fundamentals of Acoustics. Fundamentals of Acoustics, 4th Edition, by Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, James V. Sanders,."

Loftus, Geoffrey R (2002): "Loftus94." In: , **1**(4), pp. 1–15.

Lösler, Stefan and Franz Zotter (2015): "Comprehensive Radial Filter Design for Practical higher-order Ambisonic Recording." In: *Fortschritte der Akustik – DAGA 2015*, (1), pp. 452–455.

Meyer, Jens and Gary Elko (2002): "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield." In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, **2**, pp. 1781–1784.

Meyer, Jens; et al. (2008): "Handling spatial aliasing in spherical array applications." In: *2008 Hands-free Speech Communication and Microphone Arrays, Proceedings, HSCMA 2008*, **57**(2), pp. 1–4.

Rafaely, Boaz (2003): "Plane Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution." In: *ISVR Technical Memorandum 910*, (May).

Rafaely, Boaz (2005): "Analysis and design of spherical microphone arrays." In: *IEEE Transactions on Speech and Audio Processing*, **13**(1), pp. 135–143.

Rafaely, Boaz (2015): *Springer Topics in Signal Processing Springer Topics in Signal Processing*.

Rafaely, Boaz; Barak Weiss; and Eitan Bachmat (2007): "Spatial aliasing in spherical microphone arrays." In: *IEEE Transactions on Signal Processing*, **55**(3), pp. 1003–1010.

Rayleigh, Lord (1907): "XII. On our perception of sound direction." In: *Philosophical Magazine Series 6*, **13**(74), pp. 214–232.

Schörkhuber, Christian; Markus Zaunschirm; and Robert Holdrich (2018): "Binaural rendering of Ambisonic signals via magnitude least squares." In: *Proceedings of DAGA 2018*, (4), pp. 339–342.

Shannon, C.E. (1998): "Communication In The Presence Of Noise (Republished)." In: *Proceedings of the IEEE*, **86**(2), pp. 447–457.

Stade, Philipp; Benjamin Bernschütz; and Maximilian Rühl (2012): "A Spatial

Audio Impulse Response Compilation Captured at the WDR Broadcast Studios." In: *27th Tonmeistertagung - VDT International Convention*, pp. 551—-567.

Williams, Earl G. (1999): *Fourier Acoustics*. London: Academic Press.

Zaunschirm, Markus; Christian Schörkhuber; and Robert Höldrich (2018): "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint." In: *The Journal of the Acoustical Society of America*, **143**(6), pp. 3616–3627.

Zotter, Franz (2009a): "Analysis and synthesis of sound-radiation with spherical arrays." In: *IEM, Univ. Musik u. darstellende Kunst Graz*, (June), p. 192.

Zotter, Franz (2009b): "Sampling Strategies for Acoustic Holography / Holophony on the Sphere Sampling Characterization." In: *Proceedings of the NAG-DAGA 2009 International Conference on Acoustics*, **0**(2), pp. 1107–1110.

Zotter, Franz and Matthias Frank (2019): *Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*.

A

Appendix

A.1 Radial Filter

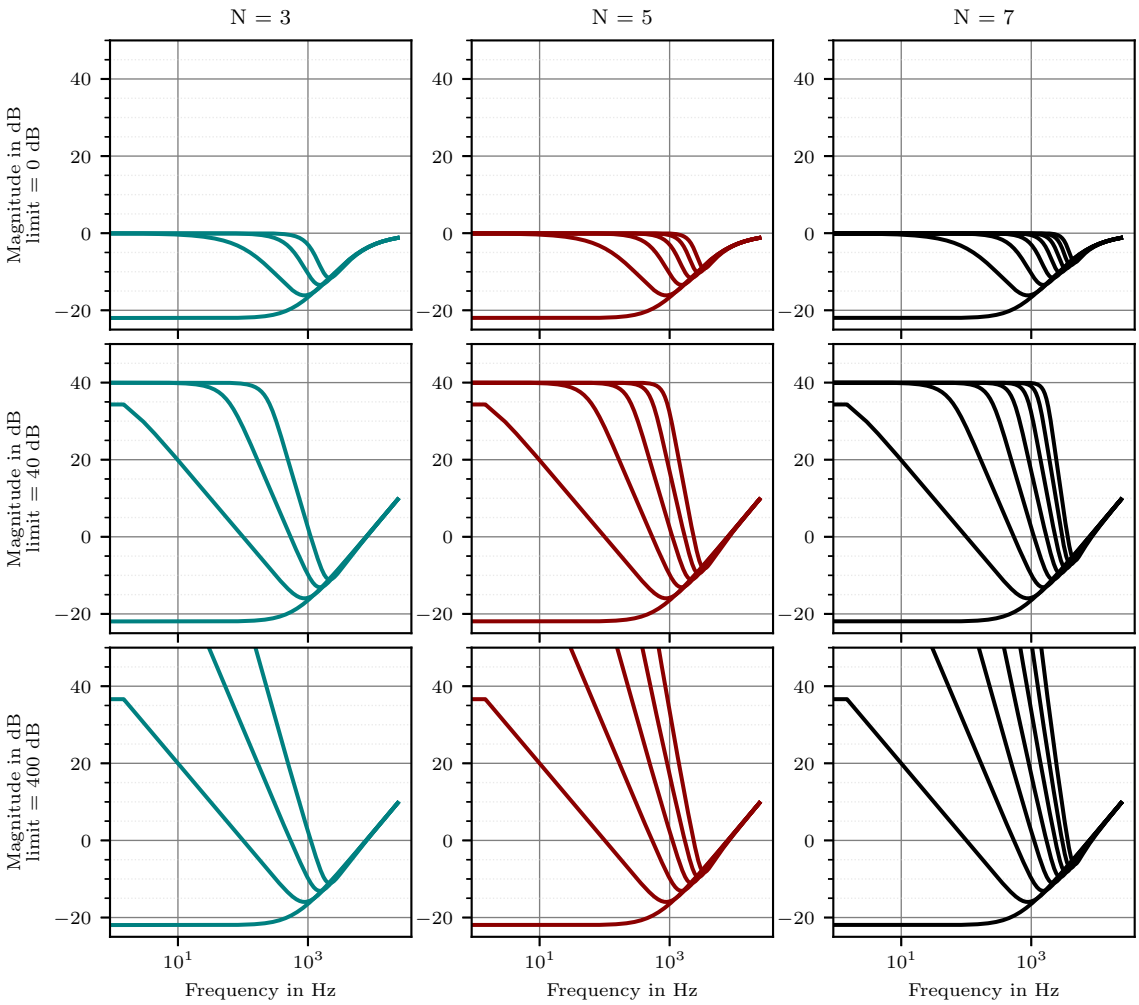

Figure A.1: 3rd, 5th and 7th order rigid-sphere radial filters with arctan soft-limiting at 0 dB, 40 dB and 400 dB (nearly unlimited). For the rendering of the listening experiments, 0 dB limited filters were used.

A.2 Comparison of Algorithm Improvements

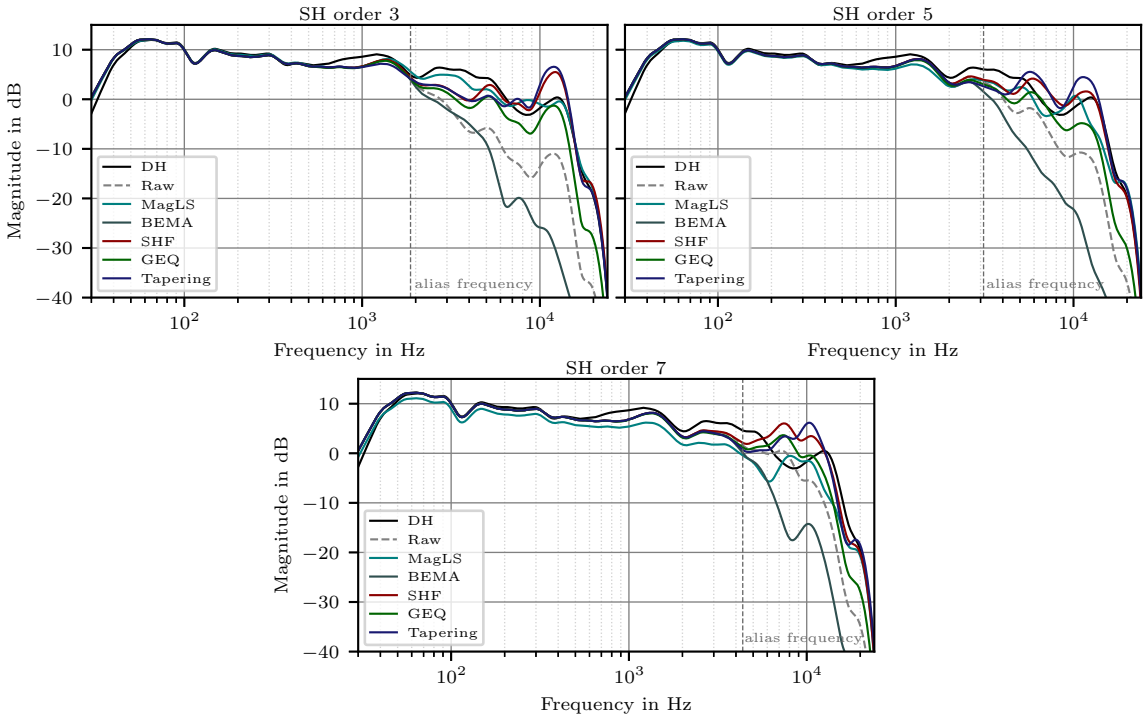

Figure A.2: Comparsion of 0° BRTFs to the left ear resulting from 5th order binaural array renderings and dummy head measurements in the LBS. The array rendering were processed with each of the proposed improvement algorithms.

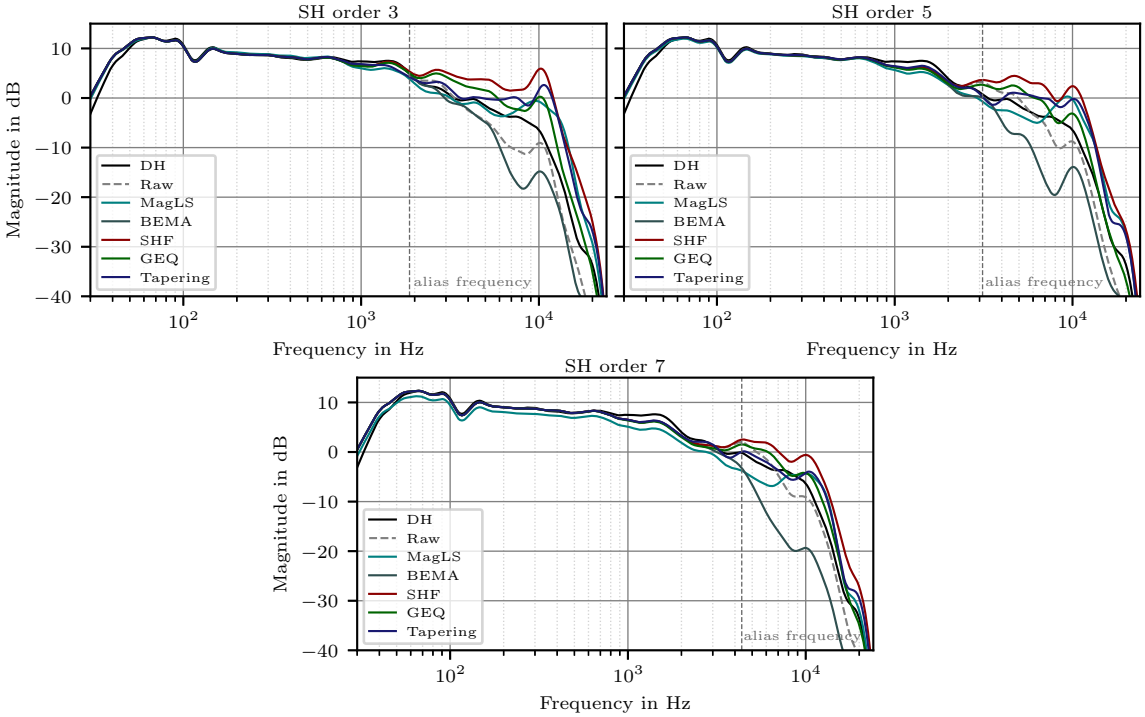

Figure A.3: Comparsion of 0° BRTFs to the left ear resulting from 5th order binaural array renderings and dummy head measurements in the LBS. The array rendering were processed with each of the proposed improvement algorithms.

A.3 Results of the Listening Experiment

A.3.1 Overall Descriptive Values

Table A.1: Statistical descriptive values mean, median, standard deviation (SD), as well as upper and lower standard 95 % confidence intervals, for all conditions divided in SBS (left), and LBS (right).

Condition SBS	Mean	Median	SD	CI-low	CI-high	Condition LBS	Mean	Median	SD	CI-low	CI-high
BEMA-N3-0	20.17	20.50	13.26	12.67	27.67	BEMA-N3-0	23.08	23.00	15.35	14.40	31.77
MagLS-N3-0	76.25	73.00	17.80	66.18	86.32	MagLS-N3-0	72.08	74.00	22.88	59.14	85.03
Tapering-N3-0	79.00	77.50	12.07	72.17	85.83	Tapering-N3-0	68.58	68.00	14.86	60.18	76.99
SHF-N3-0	75.42	73.00	14.33	67.31	83.52	SHF-N3-0	75.67	74.50	15.84	66.70	84.63
GEQ-N3-0	75.75	74.00	13.78	67.95	83.55	GEQ-N3-0	76.58	83.50	17.15	66.88	86.29
Raw-N3-0	48.75	49.00	18.15	38.48	59.02	Raw-N3-0	50.75	56.00	26.87	35.54	65.96
BEMA-N5-0	33.67	31.50	19.70	22.52	44.81	BEMA-N5-0	34.83	36.00	20.57	23.20	46.47
MagLS-N5-0	78.00	78.50	13.20	70.53	85.47	MagLS-N5-0	82.33	80.50	10.80	76.22	88.44
Tapering-N5-0	84.00	84.50	11.74	77.36	90.64	Tapering-N5-0	85.58	90.00	12.87	78.30	92.86
SHF-N5-0	81.58	80.00	8.84	76.58	86.58	SHF-N5-0	69.83	75.00	13.68	62.09	77.58
GEQ-N5-0	86.83	92.00	11.42	80.37	93.30	GEQ-N5-0	82.33	82.50	12.82	75.08	89.58
Raw-N5-0	68.75	74.50	20.46	57.17	80.33	Raw-N5-0	65.42	70.00	14.86	57.01	73.82
BEMA-N7-0	39.33	32.00	22.60	26.54	52.12	BEMA-N7-0	40.67	36.50	20.64	28.99	52.35
MagLS-N7-0	81.92	86.50	15.11	73.37	90.46	MagLS-N7-0	85.42	88.00	11.69	78.80	92.03
Tapering-N7-0	70.58	72.00	18.75	59.98	81.19	Tapering-N7-0	70.50	75.50	21.74	58.20	82.80
SHF-N7-0	75.83	73.00	12.59	68.71	82.96	SHF-N7-0	77.58	77.00	15.48	68.83	86.34
GEQ-N7-0	82.83	84.50	9.71	77.34	88.33	GEQ-N7-0	77.42	75.50	11.81	70.73	84.10
Raw-N7-0	70.08	73.00	17.24	60.33	79.84	Raw-N7-0	77.58	82.00	16.58	68.20	86.97
BEMA-N3-90	33.58	26.00	22.76	20.71	46.46	BEMA-N3-90	32.00	29.50	18.63	21.46	42.54
MagLS-N3-90	75.17	76.00	9.58	69.75	80.59	MagLS-N3-90	66.08	69.50	14.98	57.61	74.56
Tapering-N3-90	65.58	67.50	19.45	54.58	76.59	Tapering-N3-90	81.83	83.50	14.90	73.40	90.27
SHF-N3-90	62.67	70.00	20.65	50.99	74.35	SHF-N3-90	64.83	56.00	19.83	53.61	76.05
GEQ-N3-90	78.92	78.50	19.78	67.72	90.11	GEQ-N3-90	66.50	68.00	21.26	54.47	78.53
Raw-N3-90	49.33	49.00	17.06	39.68	58.99	Raw-N3-90	45.75	48.00	16.39	36.48	55.02
BEMA-N5-90	36.25	36.00	19.76	25.07	47.43	BEMA-N5-90	36.42	33.50	18.23	26.10	46.73
MagLS-N5-90	72.17	74.00	16.70	62.72	81.62	MagLS-N5-90	82.17	86.50	14.03	74.23	90.11
Tapering-N5-90	83.17	82.00	12.07	76.34	89.99	Tapering-N5-90	85.08	90.00	15.34	76.41	93.76
SHF-N5-90	74.75	77.00	18.60	64.23	85.27	SHF-N5-90	67.75	75.00	22.75	54.88	80.62
GEQ-N5-90	73.83	75.00	16.94	64.25	83.42	GEQ-N5-90	79.08	80.50	19.39	68.11	90.06
Raw-N5-90	69.83	73.50	18.00	59.65	80.02	Raw-N5-90	57.58	59.00	16.61	48.19	66.98
BEMA-N7-90	39.58	38.00	21.19	27.60	51.57	BEMA-N7-90	38.08	44.00	22.63	25.28	50.89
MagLS-N7-90	75.92	79.00	18.64	65.37	86.46	MagLS-N7-90	73.92	74.50	18.85	63.25	84.58
Tapering-N7-90	84.50	83.50	10.74	78.42	90.58	Tapering-N7-90	85.33	91.50	15.90	76.34	94.33
SHF-N7-90	82.75	86.50	15.29	74.10	91.40	SHF-N7-90	76.92	76.50	15.02	68.42	85.42
GEQ-N7-90	83.83	88.00	11.23	77.48	90.19	GEQ-N7-90	80.00	86.00	17.15	70.30	89.70
Raw-N7-90	67.25	71.00	18.04	57.05	77.45	Raw-N7-90	74.25	79.50	19.32	63.32	85.18

A.3.2 Nested ANOVA Results

Table A.2: Results of the three-way repeated measures ANOVA for the BEMA ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Order	2	18.546	.858	.537	<.001	<.001
Direction	1	6.221	1.000	.280	.024	.024
Room	1	.243	1.000	.015	.629	.629
Order×Direction	2	2.047	.964	.113	.146	.148
Order×Room	2	.503	.912	.030	.610	.593
Direction×Room	1	.001	1.000	<.001	.971	.971
Order×Direction×Room	2	.726	.784	.043	.492	.462

ϵ_{GG}: Greenhouse-Geisser epsilons
p_{GG}: Greenhouse-Geisser corrected p-values.
Statistical significance at 5 % level are indicated by asterisks. This holds for all ANOVA result tables in this appendix.

Table A.3: Results of the three-way repeated measures ANOVA for the MagLS ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Order	2	3.903	.983	.262	.035*	.036*
Direction	1	6.806	1	.382	.024*	.024*
Room	1	.040	1	.004	.845	.845
Order×Direction	2	.509	.980	.044	.608	.605
Order×Room	2	2.496	.698	.185	.105	.127
Direction×Room	1	.371	1	.033	.555	.555
Order×Direction×Room	2	.651	.652	.056	.531	.473

Table A.4: Results of the three-way repeated measures ANOVA for the Tapering ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Order	2	12.595	.614	.534	< .001*	.002*
Direction	1	5.012	1.000	.313	.047*	.047*
Room	1	.720	1.000	.061	.414	.414
Order×Direction	2	4.296	.844	.281	.027*	.035*
Order×Room	2	.105	.863	.009	.901	.874
Direction×Room	1	5.963	1.000	.352	.033*	.033*
Order×Direction×Room	2	4.849	.955	.306	.018*	.020*

Table A.5: Results of the three-way repeated measures ANOVA for the SHF ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Order	2	6.518	.929	.372	.006*	.007*
Direction	1	2.750	1.000	.200	.125	.125
Room	1	3.206	1.000	.226	.101	.101
Order×Direction	2	4.430	.961	.287	.024*	.026*
Order×Room	2	2.351	.954	.176	.119	.122
Direction×Room	2	.889	.599	.075	.425	.382
Order×Direction×Room	2	2.114	.731	.117	.137	.153

Table A.6: Results of the three-way repeated measures ANOVA for the GEQ ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	F	ϵ_{GG}	η_p^2	p	p_{GG}
Order	2	4.423	.819	.287	.024*	.034*
Direction	1	1.313	1.000	.107	.276	.276
Room	1	3.004	1.000	.214	.111	.111
Order×Direction	2	2.724	.853	.198	.088	.098
Order×Room	2	1.018	.665	.085	.378	.354
Direction×Room	1	.021	1.000	.002	.888	.888
Order×Direction×Room	2	2.424	.845	.181	.112	.122

Table A.7: Results of the three-way repeated measures ANOVA for the Raw ratings with the within-subject factors order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	*F*	ϵ_{GG}	η_p^2	*p*	p_{GG}
Order	2	23.411	.971	.680	$< .001$	$< .001^*$
Direction	1	.922	1	.077	.358	.358
Room	1	.038	1	.003	.849	.849
Order:Direction	2	.016	.933	.001	.984	.979
Order:Room	2	5.751	.678	.343	$.010^*$	$.022^*$
Direction:Room	1	1.881	1	.146	.198	.198
Order:Direction:Room	2	.271	.811	.024	.765	.720

Table A.8: Results of the four-way repeated measures ANOVA with the within-subject factors algorithm (MagLS, Tapering, SHF, GEQ) order (3, 5, 7), direction (0° , 90°) and room (SBS, LBS)

Effect	*df*	*F*	ϵ_{GG}	η_p^2	*p*	p_{GG}
Algorithm	3	2.267	.702	.171	.099	.124
Order	2	14.225	.628	.564	$< .001^*$	$< .001^*$
Direction	1	2.549	1.000	.188	.139	.139
Room	1	1.567	1.000	.125	.237	.237
Algorithm×Order	6	2.223	.608	.168	.052	.089
Algorithm×Direction	3	3.631	.909	.248	$.023^*$	$.027^*$
Order×Direction	2	5.054	.911	.315	$.016^*$	$.019^*$
Algorithm×Room	3	1.641	.872	.130	.199	.206
Order×Room	2	.297	.944	.026	.746	.734
Direction×Room	1	.944	1.000	.079	.352	.352
Algorithm×Order×Direction	6	2.140	.699	.163	.060	.088
Algorithm×Order×Room	6	1.984	.550	.153	.080	.128
Algorithm×Direction×Room	3	1.361	.760	.110	.272	.276
Order×Direction×Room	2	1.417	.996	.114	.264	.264
Algorithm×Order×Direction×Room	6	2.293	.553	.172	.045	.089