Friedrich-Alexander-Universität Erlangen-Nürnberg

Chair of
Multimedia Communications and Signal Processing

Master Thesis

Development of a Software Tool for Ambisonics Decoder using Directional Penalty Constraints

by Florian Josef Rettelbach

June 2014

Supervisors:

Mark Poletti, PhD

Prof. Dr.-Ing. Walter Kellermann
M.Sc. Michael Bürger

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Ort, Datum	Unterschrift

Lehrstuhl für Multimediakommunikation
und Signalverarbeitung
Prof. Dr.-Ing. A. Kaup

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

Masterarbeit

für

Herrn cand. Ing. Florian Rettelbach

Development of a Software Tool for Ambisonics Decoders using Directional Penalty Constraints

Surround sound systems work well at low frequencies, but as the frequency rises the area of reproduction shrinks until it is smaller than the head, after which reproduction produces aliasing artefacts. The simplest way to reduce these artefacts is to restrict the sound energy to the two nearest speakers. A number of techniques have been developed to do this, often based on the optimisation of the "energy vector" and velocity vector. However, optimisation requires nonlinear techniques, leading to complex calculations. Furthermore the energy vector is not a property of the sound field and there is little evidence that its optimisation correlates with objective or subjective measures of spatial accuracy. Lastly, the energy vector is defined at the centre of the reproduction array, and so contains no information about the reproduction accuracy over a finite volume of space. Optimisations at the centre may therefore produce less-than-optimum reproduction over the listening area.

In [Poletti, JAES 2007] a technique was developed that produced a linear least squares solution for loudspeaker weights with a directional penalty and it was shown that this produced better reproduction accuracy than techniques based on the energy vector. In this thesis, directional penalty functions are further investigated considering the accuracy for a variety of array sizes. The performance of the resulting algorithm is then compared to other techniques where published designs are available.

The software interface for deriving frequency-dependent panning filters for optimal surround sound reproduction can be implemented either in C or in Matlab. To evaluate the accuracy of the panning functions, simulations in Matlab or listening tests in an anechoic chamber should be carried out. The actual reproduced field could also be quantified using an Eigenmike. Well-documented and well-structured software is important.

Ausgabe: 01.01.2014
Abgabe: 30.06.2014

(Prof. Dr.-Ing. W. Kellermann)

Contents

Kurzfassung **VII**

Abstract **IX**

1 Introduction **1**

2 Higher Order Ambisonics **5**

2.1 Matching in Spherical Coordinates 5

2.1.1 Spherical Sound Field Description 5

2.1.2 Plane Wave Matching in Spherical Coordinates 7

2.1.3 The Right Choice of the Modes to Match and Bounds for the Sound Field 9

2.1.4 Matrix Formulation of the Azimuthal Mode Matching Equation 14

2.1.5 Matching for Point Source Loudspeaker 14

2.1.6 Direct Solution for Loudspeaker Weightings 18

2.2 Matching in Cylindrical Coordinates 21

2.2.1 Cylindrical Sound Field Description 21

2.2.2 Plane Wave Matching in Cylindrical Coordinates 23

2.2.3 Matching for Line Source Loudspeaker 24

2.2.4 Direct Solution for Loudspeaker Weightings 26

2.3 Connection between Line and Point Sources 28

2.4 Solutions for the Loudspeaker Weightings 29

2.4.1 Unregulated Solution 29
2.4.2 Tikhonov Regularization 31
2.4.3 Limit the Panning Function Order 32

3 Weighting Function Design 35
3.1 Weightings of the ITU Layout 35
3.2 Penalty Function Design 37
3.2.1 Design 37
3.2.2 Penalty Function Analysis 40
3.2.3 From Pairwise Penalty to the Stereo Panning Laws 43
3.3 Frequency Dependent Design 46
3.3.1 Robust Panning for given Reproduction Setups 47
3.3.2 Interpolation between the required Weighting Factors 48
3.3.3 Reducing the Impulse Response Coefficients 50

4 Audio Control Interface 52
4.1 Graphical User Interface 52
4.2 Test Signal Creation 55

5 Listening Test 57
5.1 Room Description and Experimental Setup 57
5.2 Implementation and Test Items 60
5.3 Results 62

6 Conclusions and Future Work 67

A Appendix 69
A.1 Spherical Harmonics 69
A.2 2D- and 3D- Greens Function 71
A.3 Questionnaire Listening Test 76
A.4 Notes Records of Questionnaire 81

B Abbreviations **93**

Abbreviations **93**

C Notation **94**

Notation **94**

List of Figures **98**

List of Tables **100**

Bibliography **102**

Kurzfassung

Im Rahmen dieser Arbeit werden Lösungsmöglichkeiten für gleich- und ungleichverteilte Lautsprecherplatzierungen auf Basis der Ambisonics-Theorie hergeleitet. Für ungleichförmig verteilte Systeme mit wenigen Lautsprechern nehmen die Amplituden der Ansteuersignale sehr hohe Werte an, was sich in großen Lautstärkeunterschieden innerhalb der Wiedergabeebene widerspiegelt. Aufgrund dessen werden Methoden zur frequenz- und winkelabhängigen Lautstärkeregulierung vorgestellt um ein einheitlicheres Schallbild im ganzen Wiedergaberaum zu erzeugen. Die vorgestellten Techniken werden nachfolgend mittels des ITU-Lautsprecherlayouts analysiert. Im weiteren wird eine graphische Benutzeroberfläche vorgestellt, welche die hier untersuchten Techniken zum Design eines Amibsonic-Decoders nutzt. Die entwickelten Verfahren werden abschließend mit Hilfe eines Hörtests analysiert.

Abstract

Solutions for uniform and nonuniform loudspeaker layouts are derived with the help of the Ambisonics theory. For nonuniform systems with a small amount of loudspeakers, the amplitudes of the driving functions attain enormous values, what results in differences of the sound pressure level within the reproduction plane. Considering this, methods for frequency- and angle-dependent level regulation are introduced to create a unified sound image within the whole reproduction area. The introduced techniques are hereinafter analyzed based on the ITU layout. Afterwards, a graphical user interface is introduced which allows the design of an Ambisonics decoder using the developed techniques. The methods are finally analyzed with the help of a listening test.

Chapter 1

Introduction

For the reproduction of spatial audio scenes, a variety of rendering methods are known. These methods either focus on a faithful replica of the initial sound field or exploit psychoacoustical features of the human hearing. On the basis of stereophony described already 1955 in [Sno55], Ville Pulkki developed the rendering method vector based amplitude panning (VBAP) [Pul97], which aims on creating interaural level and time differences at the ear to simulate phantom sources. This method uses features of the human hearing to create a sound image on a line between two speakers or a plane within three speakers. Another widely spread reproduction method is wave field synthesis (WFS). This technique aims on an exact synthesis of a desired sound field. Generally WFS needs a huge number of loudspeakers to work in a useful way. It is based on the idea that the sound field in a source free region is fully defined by the pressure on the surrounding surface. For the two-dimensional reproduction, the loudspeakers are placed on a circle or rectangle, for the three-dimensional case on a sphere. Depending on the number of loudspeakers artifacts arise for higher frequencies and the area in which the sound field can be synthesized correctly shrinks. A description of the theory is given in [SRA08]. In the scope of this thesis, the Ambisonics technique is used to create the sound field. The classical Ambisonics was introduced already in 1973 by the British mathematician Gerzon [Ger73] and extended by Daniel to Higher Order Ambisonics (HOA) [J. 01]. HOA describes a sound field in the origin with the help of

spherical harmonics for the three-dimensional case and with cylindrical harmonics for the 2D-reproduction. Spherical harmonics are orthogonal like sine and cosine functions but are defined on a sphere. The description can be understood as Fourier series expansion in the three dimensional space. A Fourier series expansion of a function is defined up to infinite order whereby it results in an approximation if it is limited up to a certain order. This holds also for the HOA sound field which is approximated with the help of spherical harmonics up to a certain order. The classical Ambisonics uses harmonics up to order one. For higher orders and more loudspeakers, the area for the synthesized sound field is extended and the frequency for correct reproduction is increased which will be seen later. In the scope of this thesis, the mathematical tools for the sound field description are provided for the three- and two-dimensional case. Furthermore, solutions for the loudspeaker weighting functions are derived. This solution can be achieved by using pressure matching [Pol07] or mode matching [Pol00]. A connection between both is derived. For plane waves, or sound source positions at the radius of a circular speaker array, the weightings that pan sound images around the circle are real and are called panning functions. In the scope of this thesis, methods for deriving panning functions for reproduction setups with a small number of loudspeakers are shown and analyzed by means of the 5.1 ITU layout [IR12]. The optimization of the panning functions of the ITU layout have been discussed before in [Moo09] [Pol07] [Cra03]. As the reproduction area gets smaller as the frequency rises, and at some frequency, the region is smaller than the head, a frequency-dependent panning function could reduce artifacts of the sound field for high frequencies. Since the artifacts are caused among other things by the side lobes of the panning functions, it is desirable to reduce the side lobes for high frequencies. Within this thesis an angular- and frequency-dependent weighting method is applied to optimize the synthesized sound field.

The following Chap. 2 provides the theoretical background for two- and three-dimensional Ambisonics. Starting from the mathematical description of a plane and radial wave in spherical and cylindrical coordinates, the solutions for the loudspeaker weightings are derived to synthesize the desired wave by matching both sound fields. Limitations

of the synthesized sound field are shown subsequently. Furthermore, direct solutions for uniformly spaced loudspeaker layouts are derived and examples given. Afterwards, methods are introduced to optimize the created sound field.

After defining the sound field and providing tools for optimization, Chap. 3 talks initially about a solution for the 5.1 ITU layout and reasons are given why this solution is not optimal. On basis of the introduced loudspeaker layout, methods for the regulation and optimization are developed and analyzed. The introduced techniques are based on an angular penalty for the loudspeakers positioned on a circle. Different penalty functions are introduced and applied on the layout. The resulting solutions are analyzed. A connection for one of these solutions to the classical stereophonic panning is given. As the Ambisonic theory provides bounds for the maximal frequency for which the sound field is synthesized correctly, a frequency-dependent solution is introduced.

In Chap. 4, a designing tool for a frequency-dependent Ambisonics decoder is shown. The functions of this graphical user interface are explained step by step. As the tool provides additionally opportunities for an audio demo, it covers all features to perform a listening test.

The last Chap. 5 talks about the performed listening test and the conditions of the reproduction setup.

Chapter 2

Higher Order Ambisonics

In this chapter, HOA will be introduced for spherical and cylindrical coordinate systems. Solutions for the loudspeaker weightings are derived and bounds for the resulting sound field are given. Furthermore, the influence of the number loudspeakers on the maximum frequency and maximum radius is analyzed.

2.1 Matching in Spherical Coordinates

HOA can be described in spherical coordinates, therefore this section will describe the mathematical foundation.

2.1.1 Spherical Sound Field Description

Spherical Coordinates Definition

The position in spherical coordinates with radius r, azimuthal angle ϕ and elevation angle θ is defined as

$$\mathbf{r} = r \begin{pmatrix} \cos(\phi)\sin(\theta) \\ \sin(\phi)\sin(\theta) \\ \cos(\theta) \end{pmatrix}. \tag{2.1}$$

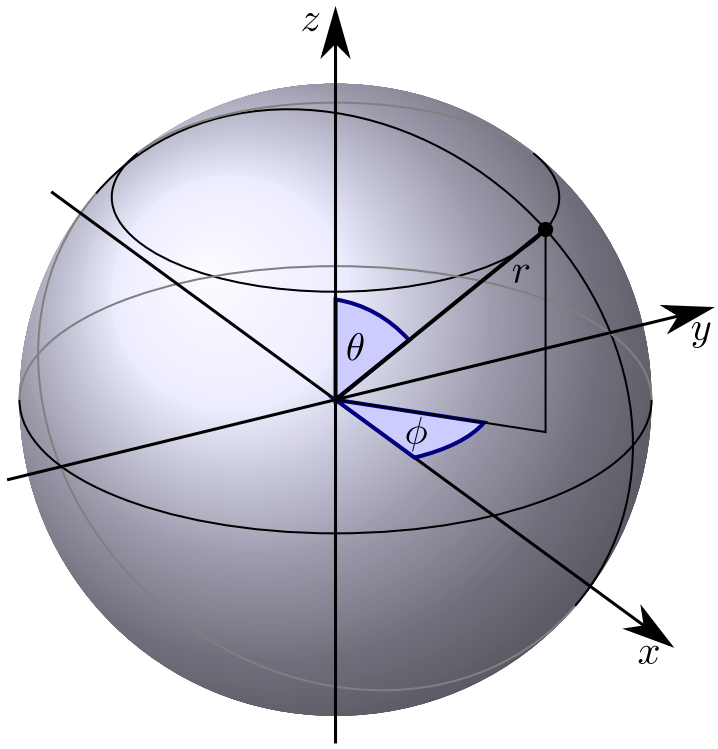

Figure 2.1: Spherical coordinates definition scheme.

The wave vector with wave number $k = \frac{\omega}{c}$, with c denoting the speed of sound and ω is the angular frequency, is defined as

$$\mathbf{k} = k \begin{pmatrix} \cos(\phi_s)\sin(\theta_s) \\ \sin(\phi_s)\sin(\theta_s) \\ \cos(\theta_s) \end{pmatrix} \tag{2.2}$$

for a wave coming from elevation angle θ_s and azimuth angle ϕ_s.

Sound Field Definition

A sound field around the origin can be generally described as in [Wil99] with

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m j_n(kr) Y_n^m(\theta, \phi), \tag{2.3}$$

where $j_n(kr)$ is the spherical Bessel function of the first kind of n-th order, in which the radius r and spatial frequency k dependency of the sound field is included. $Y_n^m(\theta, \phi)$ is the spherical Harmonic. Both are described in more detail in the Sect. A.1. A_n^m is the Ambisonic coefficient and will be explained in the following.

Ambisonic Coefficients

The Ambisonic coefficient $A_n^m(k)$ can be expressed for a plane wave as [Pol05]

$$A_n^m = 4\pi \mathrm{i}^n Y_n^m(\theta_s, \phi_s)^*, \tag{2.4}$$

where θ_s and ϕ_s is the elevation and azimuth angle of the incoming wave.
Considering a point source resp. spherical wave, the Ambisonic coefficient can be expressed as [Pol05]

$$A_n^m = \mathrm{i}k h_n(kr_s) Y_n^m(\theta_s, \phi_s)^*, \tag{2.5}$$

where $h_n(kr_s)$ is the spherical Bessel function of the third kind of n-th order. The position of the point source is defined by the radius r_s and the angles θ_s and ϕ_s in spherical coordinates.

2.1.2 Plane Wave Matching in Spherical Coordinates

Pressure and Mode Matching

A common simplification for the reproduction of sound fields is that one considers plane waves for loudspeakers and the sound sources one wants to recreate. To reproduce the sound field by the given loudspeakers, one tries to match the pressure of a plane wave at $p = 1...P$ positions $\mathbf{r}_p$ with the pressure of the plane waves produced by the loudspeakers. The pressure matching leads, depending on the number of points P, to an accurate result for the loudspeaker weightings w_l. It will be shown that the pressure matching approach can be transformed into mode matching which represents

the minimum matching conditions. First of all the pressure matching approach can be described as [Pol05] [Pol07]

$$e^{\mathbb{i}\mathbf{k}_s^{\mathbb{T}}\mathbf{r}_p} = \sum_{l=1}^{L} w_l(\theta_s, \phi_s) e^{\mathbb{i}\mathbf{k}_l^{\mathbb{T}}\mathbf{r}_p}, \tag{2.6}$$

where $\mathbf{k}_s \in \mathbb{R}^3$ is the wave vector specifying the spatial frequency components of the plane wave and $l = 1...L$ is the index for on of L loudspeakers. Making use of the spherical plane wave expansion which results from the plane wave coefficient Eq. 2.4 and the general sound field description

$$e^{\mathbb{i}\mathbf{k}_s^{\mathbb{T}}\mathbf{r}_p} = 4\pi \sum_{n=0}^{\infty} \mathbb{i}^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta_p, \phi_p) Y_n^m(\theta_s, \phi_s)^*, \tag{2.7}$$

the pressure matching approach gives

$$\begin{aligned} 4\pi \sum_{n=0}^{\infty} \mathbb{i}^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta_p, \phi_p) Y_n^m(\theta_s, \phi_s)^* = \\ = \sum_{l=1}^{L} w_l(\theta_s, \phi_s) 4\pi \sum_{n=0}^{\infty} \mathbb{i}^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta_p, \phi_p) Y_n^m(\theta_l, \phi_l)^*. \end{aligned} \tag{2.8}$$

Since the spherical Bessel functions $j_n(kr)$ are orthogonal to each other [CK12], it is just necessary to compare the respective indices n what reduces the complexity of the latter equation [WA01]

$$\begin{aligned} \mathbb{i}^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta_p, \phi_p) Y_n^m(\theta_s, \phi_s)^* = \\ = \mathbb{i}^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta_p, \phi_p) \sum_{l=1}^{L} w_l(\theta_s, \phi_s) Y_n^m(\theta_l, \phi_l)^*. \end{aligned} \tag{2.9}$$

Another orthogonality is given for the spherical Harmonics $Y_n^m(\theta_p, \phi_p)$ [CK12] and so one can compare only the terms for the particular index m

$$Y_n^m(\theta_p, \phi_p) Y_n^m(\theta_s, \phi_s)^* = Y_n^m(\theta_p, \phi_p) \sum_{l=1}^{L} w_l(\theta_s, \phi_s) Y_n^m(\theta_l, \phi_l)^*. \tag{2.10}$$

One can see that setting the plane wave expansion into the pressure matching approach results in the spherical mode matching equation

$$Y_n^m(\theta_s, \phi_s)^* = \sum_{l=1}^{L} w_l(\theta_s, \phi_s) Y_n^m(\theta_l, \phi_l)^*. \tag{2.11}$$

Simplification to Planar Reproduction

By using the definition of the spherical harmonics from Sect. A.1 and the complex conjugated property $Y_n^m(\theta, \phi)^* = Y_n^{-m}(\theta, \phi)$, one obtains

$$\Lambda_n^m P_n^{|m|}(\cos(\theta_s)) e^{-\mathrm{i}m\phi_s} = \sum_{l=1}^{L} w_l(\theta_s, \phi_s) \Lambda_n^m P_n^{|m|}(\cos(\theta_l)) e^{-\mathrm{i}m\phi_l}, \tag{2.12}$$

whereby $\Lambda_n^{-m} = \Lambda_n^m = \sqrt{\left(\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}\right)}$ and $P_n^{|m|}(\cos(\theta))$ is the $|m|$th associated Legendre function of order n. Assuming that the reproduction setup is located within a plane, the elevation angles of the loudspeakers and the source are set to $\theta_l = \theta_s = \frac{\pi}{2}$ and so $\cos(\theta_l) = \cos(\theta_s) = 0$. Eq. 2.12 reduces to the azimuthal mode matching equation

$$e^{-\mathrm{i}m\phi_s} = \sum_{l=1}^{L} w_l(\phi_s) e^{-\mathrm{i}m\phi_l}. \tag{2.13}$$

2.1.3 The Right Choice of the Modes to Match and Bounds for the Sound Field

As one can see from Eq. 2.8, one needs to match an infinite amount of modes created by the loudspeakers respectively by the sound source. An infinite amount of modes would mean an infinite amount of loudspeakers on a sphere or in the planar case on a circle creating these modes. Trapped in a sphere of loudspeakers could cause claustrophobia while going into a circle of loudspeakers needs also some physical effort not to mention

the funding and feasibility. What are the limitations for the sound field if only modes up to order N are used, i.e., the number of terms in the originally infinite sum is bounded? Ward and Abhayapala gave a rule of thumb [WA01] for the determination of N which is a trade off between the area within radius r_{max} and the maximum frequency resp. wave number k_{max}, one wants to produce. For a given frequency, the normalized error is around 4% within radius r_{max} and the maximum order defined as [WA01]

$$N_{max} = \lceil k_{max} r_{max} \rceil \,. \tag{2.14}$$

The maximum frequency f_{max}, which is possible to reproduce, is then given by

$$f_{max} = \frac{cN}{2\pi r}, \tag{2.15}$$

also known as spatial Nyquist frequency. The definition for the wave number $k = 2\pi f/c$ is used here. The maximum radius r_{max} can be calculated with

$$r_{max} = \frac{cN}{2\pi f}. \tag{2.16}$$

Further bounds for N are given in [KSAJ07]. It is then possible to reproduce the modes within a sphere with [WA01]

$$L_{max} = (N_{max} + 1)^2. \tag{2.17}$$

Fig. 2.2 suggests the modes, which are used to match the sound field. The amount of modes, which are situated within the triangle, is $(N_{max}+1)^2$. For the reproduction in a plane, only a few modes contribute noticeably to the sound field amplitude, that is only the modes with $n = |m|$. The modes marked with a $\circ$ are zero for $\theta = \pi/2$, whereas the modes marked with a $\square$ are dropped so that the remaining terms marked with a $\blacksquare$ are used for the match. This simplification is called the sectorial approximation [PBA12], mentioned already in [AS08b] [AS08a], and reduces the effort for the mode matching to $-M \leq m \leq M$ with $n = |m|$ where the rule of thumb $M = \lceil k_{max} r_{max} \rceil$ is the same as for a sphere, i.e., maximum frequency and radius are also given by Eq. 2.15 and

Eq. 2.16. The required number of loudspeakers is given by [WA01]

$$L \geq 2M + 1. \tag{2.18}$$

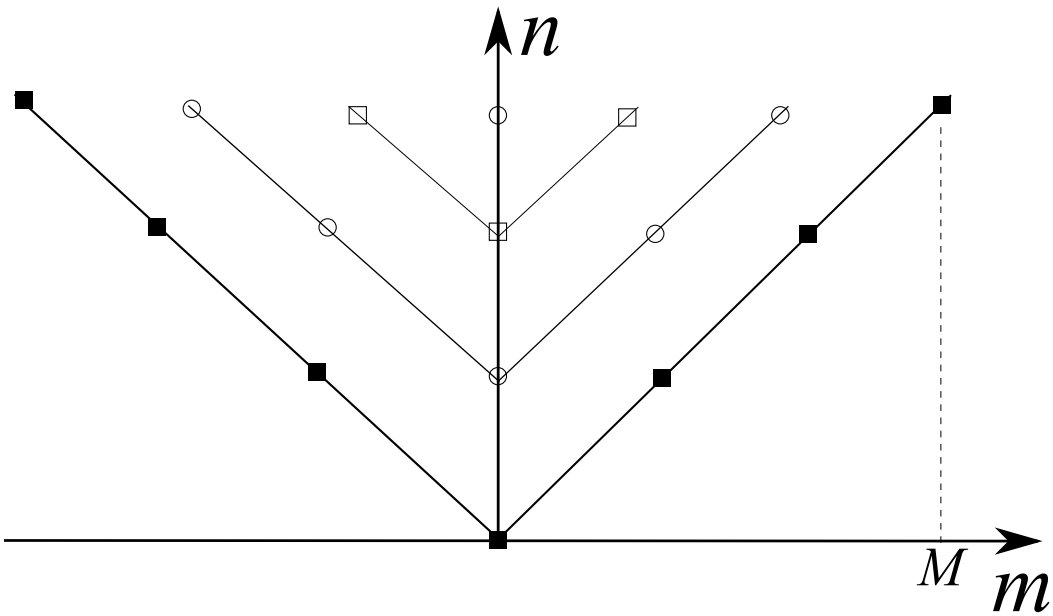

Figure 2.2: Modes to match given by the sectorial approximation marked with ■. Modes which are zero for $\theta = \pi/2$ marked with ∘. Left out modes marked with □ (adapted from [PBA12]).

In Sect. A.1, a brief analysis of the spherical harmonics up to order $N = 3$ is given. The sectorial approximation assumes that it is more useful for reducing the error to match modes of higher order marked with a ■ than the ones marked with a □.

Another limit for the number of loudspeakers, which will be used in this thesis, is shown in [SS06]. For an average error of less than −40dB, it should be higher than [SS06]

$$L \geq 2M + 3. \tag{2.19}$$

In this section so far, bounds for synthesized sound field have been given. To get an impression how the number of loudspeakers influences the sound field, Fig. 2.3 shows a sound field with 29 loudspeakers in the horizontal plane synthesizing a plane wave with a frequency of $f_{pw} = 1000$ coming from 0°. The maximum radius for the given frequency can be calculated as $r_{max} \approx 0.76$m as indicated in the plot. It is computed by using Eq. 2.16 and setting the maximum order to $M = L - 1/2 = 14$.

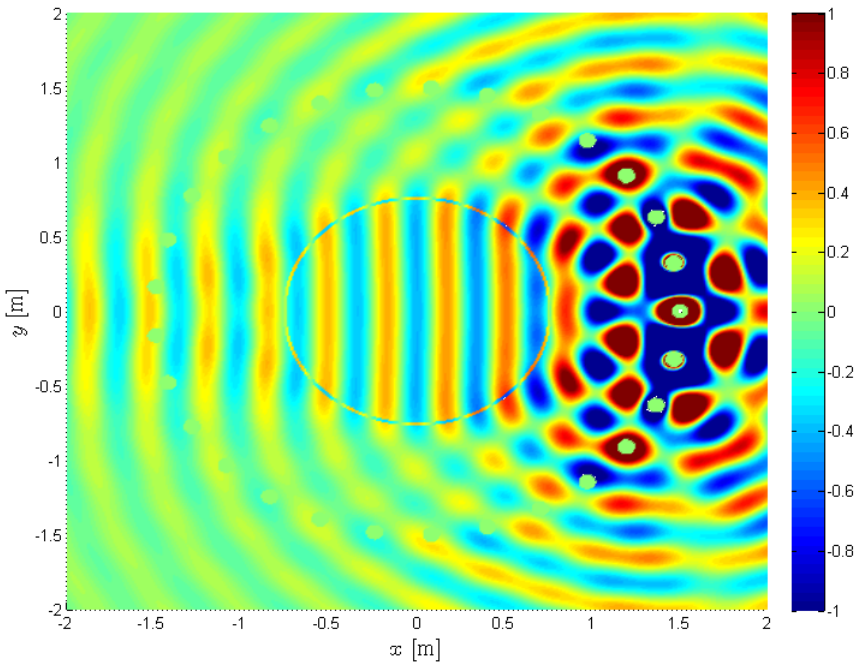

Figure 2.3: Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 29 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 1000$Hz.

If one increases the number of loudspeakers, the correctly synthesized area increases as shown in Fig. 2.4. The maximum radius $r_{max} \approx 1.47$m reaches in this case almost the radius of the circular loudspeaker array with 55 spherical point sources.

The influence of the frequency on the sound field is depicted in Fig. 2.5. The radius for the correctly reproduced sound field shrinks around the origin.

As the limits for the mode matching order are defined, the next section deals with the matrix formulation of Eq. 2.13.

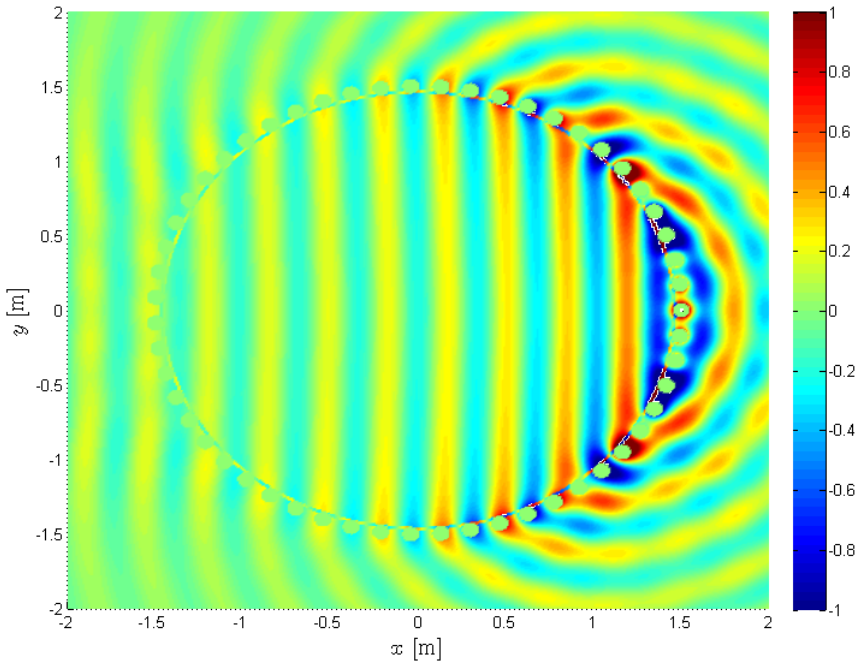

Figure 2.4: Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 55 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 1000$Hz.

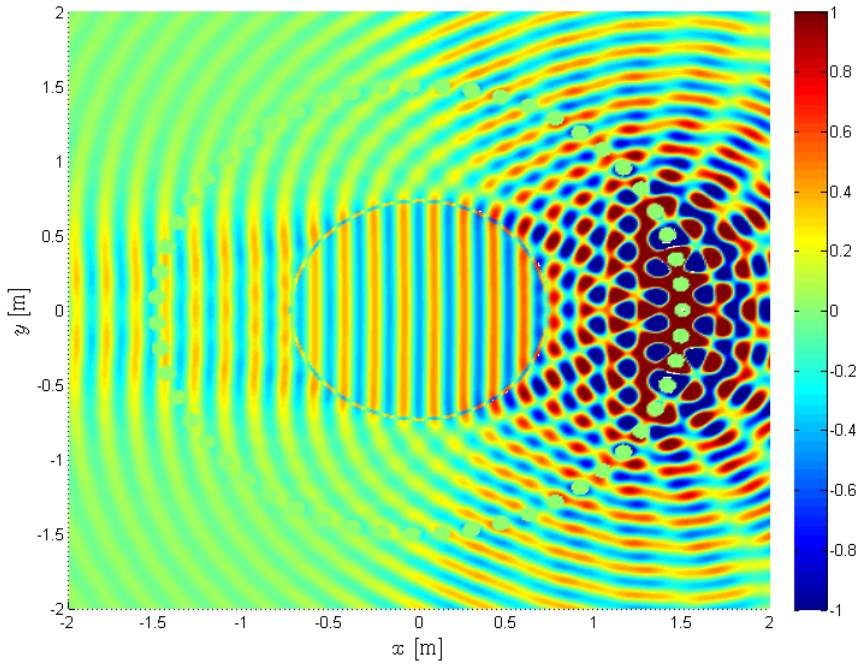

Figure 2.5: Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 55 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 2000$Hz.

2.1.4 Matrix Formulation of the Azimuthal Mode Matching Equation

Through the determination of m, the azimuthal mode matching Eq. 2.13 can be written in matrix form as

$$\begin{pmatrix} e^{-\mathbb{i}(-M)\phi_s} \\ \vdots \\ e^{-\mathbb{i}M\phi_s} \end{pmatrix} = \begin{pmatrix} e^{-\mathbb{i}(-M)\phi_1} & \dots & e^{-\mathbb{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ e^{-\mathbb{i}M\phi_1} & \dots & e^{-\mathbb{i}M\phi_L} \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_L \end{pmatrix} \tag{2.20}$$

or in compact form

$$\mathbf{p} = \mathbf{H}\mathbf{w}. \tag{2.21}$$

$\mathbf{p}$ is a vector of size $(2M+1)$ representing the modes, which has to be matched. $\mathbf{H}$ is a matrix of size $(2M+1) \times L$ including the matching conditions for a plane wave. $\mathbf{w}$ is a vector of size L, which represents the loudspeaker panning factors.

2.1.5 Matching for Point Source Loudspeaker

Up to this point, the goal was to create a plane wave with the help of several plane waves coming from certain directions. Normally, it is not a good approximation to equal the sound coming from a loudspeaker with a plane wave. Furthermore, it is often desired to synthesize the sound field created by a point source. The following equation generalizes the approach for matching a sound field at the origin with the help of point sources [BK10]

$$\sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m j_n(kr) Y_n^m(\theta, \phi) = \sum_{l=1}^{L} w_l(\phi_s) G_{3D}(\|\mathbf{r} - \mathbf{r}_l\|). \tag{2.22}$$

Inserting the spherical expansion for the 3D Green's function G_{3D}, i.e., the general expression for a sound field Eq. 2.3 with the Ambisonic coefficient for a point source

Eq. 2.5, one gets

$$\begin{aligned} \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_n^m j_n(kr) Y_n^m(\theta,\phi) = \\ \sum_{l=1}^{L} w_l(\phi_s) \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \mathbb{i} k h_n(kr_l) Y_n^m(\theta_l,\phi_l)^* j_n(kr) Y_n^m(\theta,\phi). \end{aligned} \tag{2.23}$$

It is then again just necessary to compare the terms with same index n because of orthogonality and the latter equation results in

$$\begin{aligned} j_n(kr) \sum_{m=-n}^{n} A_n^m Y_n^m(\theta,\phi) = \\ \mathbb{i} k j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta,\phi) \sum_{l=1}^{L} w_l(\phi_s) h_n(kr_l) Y_n^m(\theta_l,\phi_l)^*. \end{aligned} \tag{2.24}$$

The same processing is done with the spherical harmonics and its orthogonality property to reduce the sum to a singular term comparison

$$A_n^m Y_n^m(\theta,\phi) = \mathbb{i} k Y_n^m(\theta,\phi) \sum_{l=1}^{L} w_l(\phi_s) h_n(kr_l) Y_n^m(\theta_l,\phi_l)^*. \tag{2.25}$$

Considering that all point sources are on a sphere with radius r_l, one can formulate a general spherical matching expression

$$A_n^m = \mathbb{i} k h_n(kr_l) \sum_{l=1}^{L} w_l(\phi_s) Y_n^m(\theta_l,\phi_l)^*. \tag{2.26}$$

Matching a Point Source

Plugging in the Ambisonic coefficient for a point source Eq. 2.5 into the general spherical matching expression Eq. 2.26 leads to

$$\mathbb{i} k h_n(kr_s) Y_n^m(\theta_s,\phi_s)^* = \mathbb{i} k h_n(kr_l) \sum_{l=1}^{L} w_l(\phi_s) Y_n^m(\theta_l,\phi_l)^*. \tag{2.27}$$

and so to

$$Y_n^m(\theta_s, \phi_s)^* = \frac{h_n(kr_l)}{h_n(kr_s)} \sum_{l=1}^{L} w_l(\phi_s) Y_n^m(\theta_l, \phi_l)^*. \tag{2.28}$$

Making use of the spherical harmonic definition, the latter gives

$$\Lambda_n^m P_n^{|m|}(\cos(\theta_s)) e^{-\mathrm{i}m\phi_s} = \frac{h_n(kr_l)}{h_n(kr_s)} \sum_{l=1}^{L} w_l(\phi_s) \Lambda_n^m P_n^{|m|}(\cos(\theta_l)) e^{-\mathrm{i}m\phi_l}. \tag{2.29}$$

As this thesis concentrates on the reproduction of sound in a plane, one can simplify the given equation by $\theta_l = \theta_s = \frac{\pi}{2}$, so $\cos(\theta_l) = \cos(\theta_s) = 0$ and cancel $P_n^{|m|}(0)$ resp. Λ_n^m on both sides out. Furthermore, the simplifications regarding the sectorial approximation as shown in Fig. 2.2 are valid. The latter considerations lead to

$$e^{-\mathrm{i}m\phi_s} = \frac{h_{|m|}(kr_l)}{h_{|m|}(kr_s)} \sum_{l=1}^{L} w_l(\phi_s) e^{-\mathrm{i}m\phi_l}. \tag{2.30}$$

In matrix form, the propagation matrix $\mathbf{H}$ can be separated in

$$\mathbf{H} = \mathbf{AB}, \tag{2.31}$$

with the weighting matrix

$$\mathbf{A}_{sp}^{ps} = \begin{pmatrix} \frac{h_M(kR_l)}{h_M(kR_s)} & 0 & \cdots & & 0 \\ 0 & \ddots & & \ddots & \\ \vdots & & \frac{h_0(kR_l)}{h_0(kR_s)} & & \vdots \\ & \ddots & & \ddots & 0 \\ 0 & & \cdots & 0 & \frac{h_M(kR_l)}{h_M(kR_s)} \end{pmatrix} \tag{2.32}$$

and the phase term matrix

$$\mathbf{B} = \begin{pmatrix} e^{-\mathrm{i}(-M)\phi_1} & \cdots & e^{-\mathrm{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ e^{-\mathrm{i}M\phi_1} & \cdots & e^{-\mathrm{i}M\phi_L} \end{pmatrix}. \tag{2.33}$$

Matching a Plane Wave

Using the plane wave Ambisonic coefficient Eq. 2.4 at the general spherical matching expression Eq. 2.26 results in

$$Y_n^m(\theta_s, \phi_s)^* = \frac{\mathbb{i}^{1-n}k}{4\pi} h_n(kr_l) \sum_{l=1}^{L} w_l(\phi_s) Y_n^m(\theta_l, \phi_l)^*. \tag{2.34}$$

By using the definition of the spherical harmonic, Eq. 2.34 can be expressed as

$$\Lambda_n^m P_n^{|m|}(\cos(\theta_s)) e^{-\mathbb{i}m\phi_s} = \frac{\mathbb{i}^{1-n}k}{4\pi} h_n(kr_l) \sum_{l=1}^{L} w_l(\phi_s) \Lambda_n^m P_n^{|m|}(\cos(\theta_l)) e^{-\mathbb{i}m\phi_l}. \tag{2.35}$$

Using the same simplification as for the point source matching, one can formulate

$$e^{-\mathbb{i}m\phi_s} = \frac{\mathbb{i}^{1-|m|}k}{4\pi} h_{|m|}(kr_l) \sum_{l=1}^{L} w_l(\phi_s) e^{-\mathbb{i}m\phi_l}. \tag{2.36}$$

While $\mathbf{B}$ is the same as before, $\mathbf{A}$ is defined as

$$\mathbf{A}_{sp}^{pw} = \frac{\mathbb{i}k}{4\pi} \begin{pmatrix} \mathbb{i}^{-M} h_M(kR_l) & 0 & \cdots & & 0 \\ 0 & \ddots & & \ddots & \\ \vdots & & h_0(kR_l) & & \vdots \\ & \ddots & & \ddots & 0 \\ 0 & & \cdots & 0 & \mathbb{i}^{-M} h_M(kR_l) \end{pmatrix}. \tag{2.37}$$

Matrix Formulation for Different Radii

In the general spherical matching Eq. 2.26 it is assumed that all loudspeakers lie on a sphere and thus have the same radius. Such a setup is not always the case. Hence, it is not possible to separate $\mathbf{H}$ in $\mathbf{A}$ and $\mathbf{B}$. Accordingly, $\mathbf{H}$ is defined for a plane wave as

$$\mathbf{H}_{sp}^{pw} = \frac{\mathbb{i}k}{4\pi} \begin{pmatrix} \mathbb{i}^{-M} h_M(kr_1) e^{-\mathbb{i}(-M)\phi_1} & \cdots & \mathbb{i}^{-M} h_M(kr_L) e^{-\mathbb{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ \mathbb{i}^{-M} h_M(kr_1) e^{-\mathbb{i}M\phi_1} & \cdots & \mathbb{i}^{-M} h_M(kr_L) e^{-\mathbb{i}M\phi_L} \end{pmatrix} \tag{2.38}$$

and for a point source as

$$\mathbf{H}_{sp}^{ps} = \begin{pmatrix} \frac{h_M(kr_1)}{h_M(kr_s)} e^{-\mathbb{i}(-M)\phi_1} & \cdots & \frac{h_M(kr_L)}{h_M(kr_s)} e^{-\mathbb{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ \frac{h_M(kr_1)}{h_M(kr_s)} e^{-\mathbb{i}M\phi_1} & \cdots & \frac{h_M(kr_L)}{h_M(kr_s)} e^{-\mathbb{i}M\phi_L} \end{pmatrix}. \tag{2.39}$$

2.1.6 Direct Solution for Loudspeaker Weightings

In the previous sections, approaches for matching sound pressure and modes were given. The equations so far do not give a good impression how the loudspeaker weightings have to look like. To derive a direct solution for the loudspeaker driving function, one uses an infinite distribution of point sources on a sphere, which is then sampled at equally spaced angular distances. This procedure is called simple source approach [Pol05] and results in the following loudspeaker weighting

$$w_l = g_l \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{A_n^m}{\mathbb{i}kh_n(kr_l)} Y_n^m(\theta_l, \phi_l), \tag{2.40}$$

where g_l is a weighting term, which comes from finite sampling points. Using the Ambisonic coefficient for a point source Eq. 2.5 and assuming that the desired source lies also on the loudspeaker sphere, one obtains

$$w_l = g_l \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^m(\theta_s, \phi_s)^* Y_n^m(\theta_l, \phi_l). \tag{2.41}$$

Using the addition theorem of the Legendre functions [CK12], the latter equation can be expressed as

$$w_l = g_l \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\mathbf{r}_s \cdot \mathbf{r}_l), \tag{2.42}$$

where $\mathbf{r}_s \in \mathbb{R}^3$ is the vector from the origin to the position of the source and $\mathbf{r}_l$ to the loudspeaker position. The scalar product between the two vectors is $\mathbf{r}_s \cdot \mathbf{r}_l = \cos\gamma$, where γ is the angle between the two vectors. To limit the equation to the spherical angles θ_s and ϕ_s describing a variable source on a sphere, one can set $\theta_l = \pi/2$ and $\phi_s = 0$. In this case, one considers the weighting function of a loudspeaker positioned on the x-axis. Furthermore the summation limit can be defined as $N = \sqrt{L} - 1$.

$$w_l = g_l \sum_{n=0}^{\sqrt{L}-1} \frac{2n+1}{4\pi} P_n(\cos\phi_s \sin\theta_s). \tag{2.43}$$

If L is not a square number, N has to be rounded down.

For a reproduction setup with $L = 25$ transducers equally distributed on the surface of a sphere, Fig. 2.6 shows the amplitude weighting for a loudspeaker placed on the x-axis at radius r_l for different positions of the source, which has to be reproduced. Inserting the number of loudspeakers in latter equation results in

$$\sum_{n=0}^{4} \frac{2n+1}{4\pi} P_n(xy) = \frac{5}{32\pi}(63x^4y^4 + 28x^3y^3 - 42x^2y^2 - 12xy + 3), \tag{2.44}$$

whereby the relation

$$x = \cos(\phi_s), y = \sin(\theta_s) \tag{2.45}$$

is used.

In this section, mathematical definitions of the sound field in spherical coordinates have been given. Furthermore, the connection between pressure and mode matching has been derived. On the basis of mode matching, the sound field of the loudspeakers and the sound image has been matched for plane and radial waves, and formulated in matrix equations. Bounds for the synthesized sound field have been given. Furthermore, a

direct solution for the loudspeaker weightings has been derived. The following section deals with the sound field definition in cylindrical coordinates and derives again the connection between pressure and mode matching.

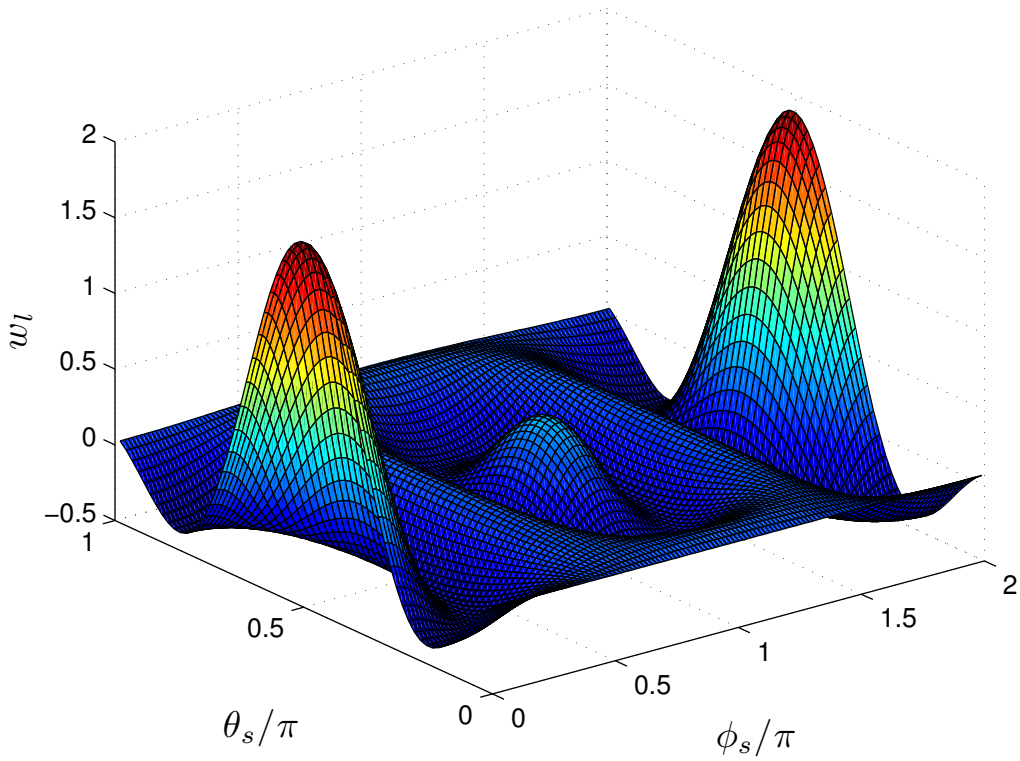

Figure 2.6: Panning function on a sphere depending on azimuthal angle ϕ_s, elevation angle θ_s of the source, for $L = 25$ loudspeaker.

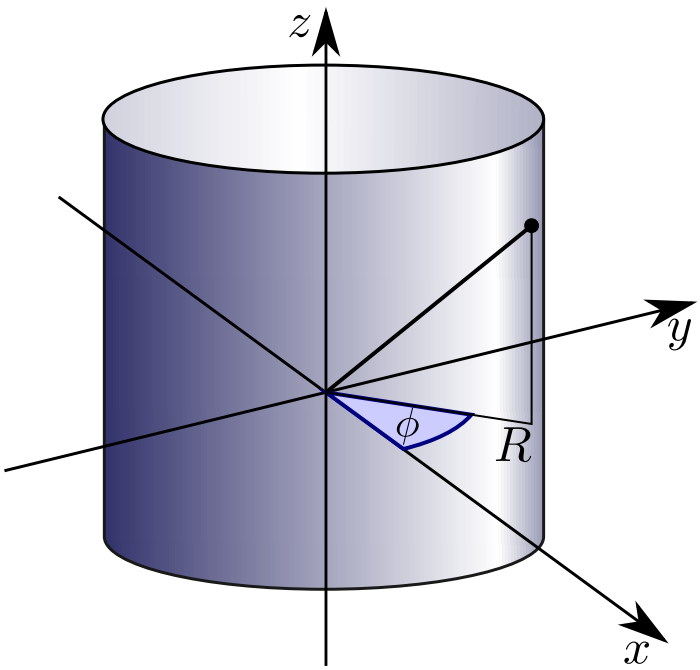

Figure 2.7: Cylindrical coordinates definition scheme.

2.2 Matching in Cylindrical Coordinates

HOA can be described in spherical coordinates, therefore this section will describe the mathematical foundation.

2.2.1 Cylindrical Sound Field Description

Cylindrical Coordinates Definition

The position in cylindrical coordinates with radius R and azimuthal angle ϕ is defined as

$$\mathbf{r} = R \begin{pmatrix} \cos(\phi) \\ \sin(\phi) \end{pmatrix}, \tag{2.46}$$

whereas z-invariance is assumed.

The wave vector $\mathbf{k}$ with wave number $k = \frac{\omega}{c}$ is defined as

$$\mathbf{k} = k \begin{pmatrix} \cos(\phi_s) \\ \sin(\phi_s) \\ 0 \end{pmatrix}, \tag{2.47}$$

where it is assumed that the z component of k is 0.

Sound Field Definition

A z-invariant wave in cylindrical coordinates can be generally expressed as [WAM09]

$$p(R, \phi, k) = \sum_{m=-\infty}^{\infty} A_m J_m(kR) e^{\mathrm{i}m\phi}, \tag{2.48}$$

where $J_m(kR)$ is the Bessel function of mth order, which is dependent on cylindrical radius R and spatial frequency k.

Ambisonic Coefficients

A_m stands for the cylindrical Ambisonic coefficient and is defined as [WAM09]

$$A_m = \mathrm{i}^m e^{-\mathrm{i}m\phi_s} \tag{2.49}$$

for a plane wave coming from the azimuthal incident angle ϕ_s.

Considering a line source or cylindrical wave, the coefficient is defined as [WAM09]

$$A_m = \frac{\mathrm{i}}{4} H_m(kR_s) e^{-\mathrm{i}m\phi_s}. \tag{2.50}$$

$H_m(kR_s)$ is the Hankel function of the first kind of m-th order. The position of the line source is given by its radius R_s and its angle ϕ_s.

2.2.2 Plane Wave Matching in Cylindrical Coordinates

As before in the spherical case, it is approached to create a plane wave by the superposition of plane waves coming from the loudspeaker directions. For this approach, the procedure of matching the pressure on the positions p with radius R_p and azimuthal angle ϕ_p is the same [Pol00]

$$e^{-\mathbb{i}k_s R_p \cos(\phi_p - \phi_s)} = \sum_{l=1}^{L} w_l(\phi_s) e^{-\mathbb{i}k_l R_p \cos(\phi_p - \phi_l)}. \tag{2.51}$$

The cylindrical Bessel expansion for a plane wave is defined as

$$p_{\phi_s}(R, \phi) = e^{-\mathbb{i}k_s R \cos(\phi - \phi_s)} = \sum_{m=-\infty}^{\infty} \mathbb{i}^m J_m(k_s R) e^{\mathbb{i}m(\phi - \phi_s)}, \tag{2.52}$$

which is the combination of the Ambisonic coefficient and the general cylindrical wave expression. The previous equation can then be written as

$$\sum_{m=-\infty}^{\infty} \mathbb{i}^m J_m(k_s R_p) e^{\mathbb{i}m(\phi_p - \phi_s)} = \sum_{l=1}^{L} w_l(\phi_s) \sum_{m=-\infty}^{\infty} \mathbb{i}^m J_m(k_l R_p) e^{\mathbb{i}m(\phi_p - \phi_l)}. \tag{2.53}$$

The orthogonality property is also valid for the Bessel function $J_m(k_s R_p)$ and so the per mth term comparison results in

$$\mathbb{i}^m J_m(k_s R_p) e^{\mathbb{i}m(\phi_p - \phi_s)} = \mathbb{i}^m J_m(k_l R_p) \sum_{l=1}^{L} w_l(\phi_s) e^{\mathbb{i}m(\phi_p - \phi_l)}. \tag{2.54}$$

By reducing both sides, latter term results in the azimuthal mode matching Eq. 2.13

$$e^{-\mathbb{i}m\phi_s} = \sum_{l=1}^{L} w_l(\phi_s) e^{-\mathbb{i}m\phi_l}, \tag{2.55}$$

in which just the phase terms have to be matched.

2.2.3 Matching for Line Source Loudspeaker

For the reproduction in a plane, one usually approximates the radiation pattern of the loudspeaker with a line source that is the two dimensional Green's function G_{2D}

$$p(R,\phi,k) = \sum_{l=1}^{L} w_l(\phi_s) G_{2D}(\|\mathbf{r}-\mathbf{r}_l\|) = \sum_{l=1}^{L} w_l(\phi_s) \frac{\mathrm{i}}{4} H_0(k\,\|\mathbf{r}-\mathbf{r}_l\|). \tag{2.56}$$

Trying to reproduce any wave in cylindrical coordinates, one can use the cylindrical Bessel expansion for a line source

$$\sum_{m=-\infty}^{\infty} A_m J_m(kR) e^{\mathrm{i}m\phi} = \frac{\mathrm{i}}{4} \sum_{l=1}^{L} w_l \sum_{m=-\infty}^{\infty} H_m(kR_l) J_m(kR) e^{\mathrm{i}m(\phi-\phi_l)}. \tag{2.57}$$

After using the orthogonality property of the Bessel function and reducing, one receives the general cylindrical matching expression

$$A_m = \frac{\mathrm{i}}{4} H_m(kR_l) \sum_{l=1}^{L} w_l e^{-\mathrm{i}m\phi_l}. \tag{2.58}$$

Matching a Line Source

Using the Ambisonic coefficient for a line source with radius R_s and azimuthal angle ϕ_s, one can formulate

$$\frac{\mathrm{i}}{4} H_m(kR_s) e^{-\mathrm{i}m\phi_s} = \frac{\mathrm{i}}{4} H_m(kR_l) \sum_{l=1}^{L} w_l e^{-\mathrm{i}m\phi_l}, \tag{2.59}$$

which reduces to

$$e^{-\mathrm{i}m\phi_s} = \frac{H_m(kR_l)}{H_m(kR_s)} \sum_{l=1}^{L} w_l e^{-\mathrm{i}m\phi_l}. \tag{2.60}$$

Writing the latter equation in matrix form Eq. 2.21, the weighting matrix $\mathbf{A}$ is then

$$\mathbf{A}_{cy}^{ls} = \begin{pmatrix} \frac{H_{-M}(kR_l)}{H_{-M}(kR_s)} & 0 & \cdots & & 0 \\ 0 & \ddots & & \ddots & \\ \vdots & & \frac{H_0(kR_l)}{H_0(kR_s)} & & \vdots \\ & \ddots & & \ddots & 0 \\ 0 & & \cdots & 0 & \frac{H_M(kR_l)}{H_M(kR_s)} \end{pmatrix}. \tag{2.61}$$

Matching a Plane Wave

Creating a plane wave with the help of line sources is straight forward

$$\mathbb{i}^m e^{-\mathbb{i}m\phi_s} = \frac{\mathbb{i}}{4} H_m(kR_l) \sum_{l=1}^{L} w_l e^{-\mathbb{i}m\phi_l} \tag{2.62}$$

or written as

$$e^{-\mathbb{i}m\phi_s} = \frac{\mathbb{i}^{1-m}}{4} H_m(kR_l) \sum_{l=1}^{L} w_l e^{-\mathbb{i}m\phi_l}. \tag{2.63}$$

This results in the slightly different $\mathbf{A}$:

$$\mathbf{A}_{cy}^{pw} = \begin{pmatrix} \frac{\mathbb{i}^{1+M}}{4} H_{-M}(kR_l) & 0 & \cdots & & 0 \\ 0 & \ddots & & \ddots & \\ \vdots & & \frac{\mathbb{i}}{4} H_0(kR_l) & & \vdots \\ & \ddots & & \ddots & 0 \\ 0 & & \cdots & 0 & \frac{\mathbb{i}^{1-M}}{4} H_M(kR_l) \end{pmatrix}. \tag{2.64}$$

Matrix Formulation for Different Radii

As already mentioned in the spherical case, it cannot always be assumed that all loudspeakers possess the same radius. For this case, matrix $\mathbf{H}$ is again not separable. For a plane wave, it results then in

$$\mathbf{H}_{cy}^{pw} = \begin{pmatrix} \frac{\mathrm{i}^{1+M}}{4} H_{-M}(kR_1) e^{-\mathrm{i}(-M)\phi_1} & \cdots & \frac{\mathrm{i}^{1+M}}{4} H_{-M}(kR_L) e^{-\mathrm{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ \frac{\mathrm{i}^{1-M}}{4} H_M(kR_1) e^{-\mathrm{i}M\phi_1} & \cdots & \frac{\mathrm{i}^{1-M}}{4} H_M(kR_L) e^{-\mathrm{i}M\phi_L} \end{pmatrix} \tag{2.65}$$

and for a line source in

$$\mathbf{H}_{cy}^{ls} = \begin{pmatrix} \frac{H_{-M}(kR_1)}{H_{-M}(kR_s)} e^{-\mathrm{i}(-M)\phi_1} & \cdots & \frac{H_{-M}(kR_L)}{H_{-M}(kR_s)} e^{-\mathrm{i}(-M)\phi_L} \\ \vdots & \ddots & \vdots \\ \frac{H_M(kR_1)}{H_M(kR_s)} e^{-\mathrm{i}M\phi_1} & \cdots & \frac{H_M(kR_L)}{H_M(kR_s)} e^{-\mathrm{i}M\phi_L} \end{pmatrix}. \tag{2.66}$$

2.2.4 Direct Solution for Loudspeaker Weightings

As the simple source approach Eq. 2.40 results in a direct, more or less simple interpretable solution, it would be useful to have a solution for loudspeakers placed on a circle and so just dependent on one angular parameter. The origin of this panning function lies in this case in an infinite line source distribution on a circle, which is then sampled at equally distant points. The weighting of one single loudspeaker can then be written as in [WAM09]

$$w_l(\phi_s) = \frac{1}{L} \sum_{m=-M}^{M} \frac{A_m}{\mathrm{i}\pi H_m(k_l R_l)} e^{\mathrm{i}m\phi_l}, \tag{2.67}$$

where L is the number of loudspeakers. Using the definition of the Ambisonic coefficient for a line source, assuming that the desired sound source lies on the same circle as the loudspeakers, i.e., $R_s = R_l$, having naturally the same frequency $k_s = k_l$, one receives for a loudspeaker at angle $\phi_l = 0$

$$w_l(\phi_s) = \frac{1}{L} \sum_{m=-M}^{M} e^{\mathrm{i}m\phi_s} = \frac{1}{L} e^{-\mathrm{i}M\phi_s} \sum_{m=0}^{2M} e^{\mathrm{i}m\phi_s}. \tag{2.68}$$

Using the relation for a finite geometric series, one receives

$$w_l(\phi_s) = \frac{1}{L} e^{-\mathrm{i}M\phi_s} \frac{1 - e^{\mathrm{i}(2N+1)\phi_s}}{1 - e^{\mathrm{i}\phi_s}}. \tag{2.69}$$

By using $M = (L-1)/2$ [WA01], latter equation can be simplified to

$$w_l(\phi_s) = \frac{1}{L} \frac{\sin(\frac{L}{2}\phi_s)}{\sin(\frac{1}{2}\phi_s)}, \tag{2.70}$$

which is the angular cardinal sine function. This function can only be used for an odd number of loudspeakers. For an even number of loudspeakers, the angular sampling function is defined as [Pol96]

$$w_l(\phi_s) = \frac{1}{L} \left(\frac{\sin(\frac{L-1}{2}\phi_s)}{\sin(\frac{1}{2}\phi_s)} + \cos(\frac{L}{2}\phi_s) \right). \tag{2.71}$$

To get an impression for the panning function, the weighting curve for a loudspeaker at $\phi_l = 0$ is shown in Fig. 2.8. The function possesses $L-1$ zeros, where L is the number of loudspeakers of the reproduction setup. If the desired sound source lies directly on a loudspeaker, all other loudspeakers are weighted with zero. In this case, the loudspeaker at position $\phi_l = 0$ gets exactly weighting $w_l = 1$ if the source lays on the same position, that is all other loudspeakers should produce no acoustic signal. For an even number of transducers, there is another zero necessary at the opposing site of the desired source as one can see from the curve. One has to remember that an even number of sampling points results in a loss of phase information, i.e., the gain for an additional loudspeaker is always higher for a resulting odd number of transducer. Proofs for the solution and analyses are given in [Pol00]. Further details for the angular sinc function for an even number of loudspeakers are given in [LS03].

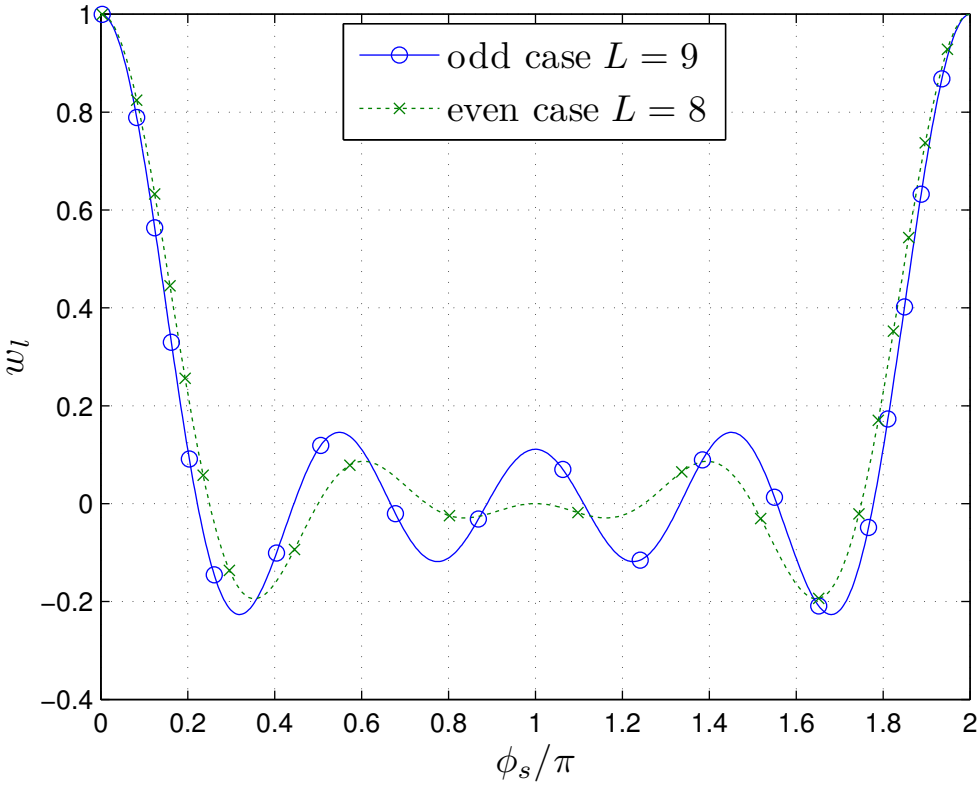

Figure 2.8: Panning function depending on the source angle ϕ_s for loudspeaker placed at $\phi_l = 0$ with a reproduction setup of equally distributed loudspeakers placed on circle.

2.3 Connection between Line and Point Sources

In the previous sections, several ways to the solution of loudspeaker weightings were given. Parallel, solutions for line respectively point sources were developed. So far it was assumed that the line sources were ideal. A brief analysis of line and point sources respectively the cylindrical and spherical Green's function is given in Sect. A.2. Line sources provide a smaller attenuation if the distance to the source increases. But even with the current technology it is some effort to produce a loudspeaker, made out of infinitely small point sources with infinite height not to mention the required space. Limited through space and technology, a loudspeaker is never an ideal line source. These limitations result in a line source like radiation. A line source with narrower

length and certain distance between the point source becomes always a spherical wave which is the far field property. Furthermore, the radiation property is frequency dependent, because the distance between the point sources is not infinitesimal small. Accordingly, it should be desired to have transducer arrays which have near field property for the whole reproduction area. Otherwise, the weighting factors should be calculated with the spherical solution to avoid additional errors. There are several producers for line arrays resp. line source alike loudspeakers which are used for home entertainment systems or PA systems. More information about line and point source and their characteristics can be found for example in [KFCS99].

2.4 Solutions for the Loudspeaker Weightings

2.4.1 Unregulated Solution

Pseudo Inverse Matrix

The solution of the equation $\mathbf{p} = \mathbf{H}\mathbf{w}$ depends on the size $(2M+1) \times L$ of the mode representing matrix $\mathbf{H}$. If $L = 2M+1$, i.e., if $\mathbf{H}$ is square, the system possesses a unique solution. The solution is then simply given by the inverse

$$\mathbf{w} = \mathbf{H}^{-1}\mathbf{p}. \tag{2.72}$$

Provided that H is nonsingular, this should be the desired case, i.e., one tries to reproduce the maximum mode order with the given number of loudspeakers. If the matrix $\mathbf{H}$ is not square, it is not possible to invert the matrix. One can use the pseudo inverse as a solution. For a system with $L > 2M+1$, that is more columns than rows, the pseudo inverse $\mathbf{H}^{+}$ is given by [PT55][Hjø11]

$$\mathbf{w} = \mathbf{H}^{+}\mathbf{p} = \mathbf{H}^{\mathbb{H}}(\mathbf{H}\mathbf{H}^{\mathbb{H}})^{-1}\mathbf{p}. \tag{2.73}$$

For a system with less columns than rows $L < 2M+1$, one minimizes the error to find

the desired solution. In this case the pseudo inverse is derived from the least squares approach [PT55][Hjø11]

$$\mathbf{w} = \mathbf{H}^{+}\mathbf{p} = (\mathbf{H}^{\mathbb{H}}\mathbf{H})^{-1}\mathbf{H}^{\mathbb{H}}\mathbf{p}. \tag{2.74}$$

The solutions for these pseudo inverses will be derived in the next sections.

Minimum Energy Solution

If $\mathbf{H}$ has full row rank, which means there are not enough modes to specify the loudspeaker weightings $L > 2M + 1$, the solution is derived with the focus on minimum weightings energy respectively loudspeaker power. The additional mode matching condition $\mathbf{p} - \mathbf{H}\mathbf{w} = 0$ completes then the Lagrange function, which is defined as

$$\Lambda(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{w}^{\mathbb{H}}\mathbf{w} + \boldsymbol{\lambda}^{\mathbb{T}}\left[\mathbf{p} - \mathbf{H}\mathbf{w}\right]. \tag{2.75}$$

To receive the minimum, one equates the derivatives with respect to $\mathbf{w}$

$$\frac{\partial \Lambda}{\partial \mathbf{w}} = \mathbf{w}^{\mathbb{H}} - \boldsymbol{\lambda}^{\mathbb{T}}\mathbf{H} \tag{2.76}$$

and $\boldsymbol{\lambda}$

$$\frac{\partial \Lambda}{\partial \boldsymbol{\lambda}} = \left[\mathbf{p} - \mathbf{H}\mathbf{w}\right]^{\mathbb{T}} \tag{2.77}$$

to zero. Afterwards one resolves Eq. 2.76 to $\mathbf{w}^{\mathbb{H}} = \boldsymbol{\lambda}^{\mathbb{T}}\mathbf{H}$ and inserts it into latter equation

$$\mathbf{p}^{\mathbb{H}} = \boldsymbol{\lambda}^{\mathbb{T}}\mathbf{H}\mathbf{H}^{\mathbb{H}}. \tag{2.78}$$

Solving for $\boldsymbol{\lambda}^{\mathbb{T}}$ and plugging into Eq. 2.76, one receives

$$\mathbf{w}^{\mathbb{H}} = \mathbf{p}^{\mathbb{H}}(\mathbf{H}\mathbf{H}^{\mathbb{H}})^{-1}\mathbf{H}. \tag{2.79}$$

Taking the Hermitian of latter equation, one obtains in the minimum energy solution Eq. 2.73 of the Moore-Penrose pseudo inverse

$$\mathbf{w} = ((\mathbf{H}\mathbf{H}^{\mathbb{H}})^{-1}\mathbf{H})^{\mathbb{H}}\mathbf{p} = \mathbf{H}^{\mathbb{H}}((\mathbf{H}\mathbf{H}^{\mathbb{H}})^{-1})^{\mathbb{H}}\mathbf{p}. \tag{2.80}$$

Least Squares Approach

As mentioned before, the pseudo inverse for a system, which has less columns than rows, is the solution for the least squares approach

$$\epsilon = \|\mathbf{p} - \mathbf{H}\mathbf{w}\|^2 , \tag{2.81}$$

which has to be minimized. Writing the L^2 norm in matrix multiplication terms

$$\epsilon = [\mathbf{p} - \mathbf{H}\mathbf{w}]^{\mathbb{H}} [\mathbf{p} - \mathbf{H}\mathbf{w}] \tag{2.82}$$

and multiplying the brackets, one gets

$$\epsilon = \mathbf{p}^{\mathbb{H}}\mathbf{p} - \mathbf{p}^{\mathbb{H}}\mathbf{H}\mathbf{w} - \mathbf{w}^{\mathbb{H}}\mathbf{H}^{\mathbb{H}}\mathbf{p} + \mathbf{w}^{\mathbb{H}}\mathbf{H}^{\mathbb{H}}\mathbf{H}\mathbf{w}. \tag{2.83}$$

Using the differentiation rules for complex vectors or matrices [Hjø11], one receives the derivative and the minimum by equaling to zero

$$\frac{\partial \epsilon}{\partial \mathbf{w}^{\mathbb{H}}} = -\mathbf{H}^{\mathbb{H}}\mathbf{p} + \mathbf{H}^{\mathbb{H}}\mathbf{H}\mathbf{w} = 0, \tag{2.84}$$

where the solution for $\mathbf{w}$ is then given by Eq. 2.74.

2.4.2 Tikhonov Regularization

If you want to regulate the total power of the loudspeaker weightings, you can use the regularization with euclidean norm also known as Tikhonov regularization [BV10] which was already used in [Pol05] [WA01] [BK10]

$$\epsilon = \|\mathbf{p} - \mathbf{H}\mathbf{w}\|^2 + \gamma \|\mathbf{w}\|^2 . \tag{2.85}$$

Limiting the power increases the robustness of the reproduction system regarding variations of the loudspeaker positioning but does not provide much room for the control of a specific loudspeaker setup. A more analytical way to create panning functions is described in [Pol07]. Poletti uses a penalty function $\gamma_l = 0.5(1-\cos(\phi_l - \phi_s))$ to reduce the power of the loudspeakers whose sound-waves come from the opposite direction, compared to source angle. Putting the elements γ_l on the diagonal of the matrix $\mathbf{\Gamma}$, which is sometimes called Tikhonov matrix, one can write Eq. 2.85 as

$$\epsilon = \|\mathbf{p} - \mathbf{H}\mathbf{w}\|^2 + \gamma_r \|\mathbf{\Gamma}\mathbf{w}\|^2 \tag{2.86}$$

with γ_r as an additional scalar regularization factor to control the influence of $\|\mathbf{\Gamma}\mathbf{w}\|^2$ on the weighting solution. In matrix product terms one receives

$$\epsilon = \mathbf{p}^{\mathbb{H}}\mathbf{p} - \mathbf{p}^{\mathbb{H}}\mathbf{H}\mathbf{w} - \mathbf{w}^{\mathbb{H}}\mathbf{H}^{\mathbb{H}}\mathbf{p} + \mathbf{w}^{\mathbb{H}}\mathbf{H}^{\mathbb{H}}\mathbf{H}\mathbf{w} + \gamma_r \mathbf{w}^{\mathbb{H}}\mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\mathbf{w}. \tag{2.87}$$

Using again the differentiation rules for complex matrices [Hjø11] and setting to zero, one receives

$$\frac{\partial \epsilon(\mathbf{w})}{\partial \mathbf{w}^{\mathbb{H}}} = -\mathbf{H}^{\mathbb{H}}\mathbf{p} + \mathbf{H}^{\mathbb{H}}\mathbf{H}\mathbf{w} + \gamma_r \mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\mathbf{w} = 0. \tag{2.88}$$

This equation can then be solved with respect to $\mathbf{w}$ and one receives the solution

$$\mathbf{w} = \left[\mathbf{H}^{\mathbb{H}}\mathbf{H} + \gamma_r \mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\right]^{-1} \mathbf{H}^{\mathbb{H}}\mathbf{p}. \tag{2.89}$$

If the matrix $\mathbf{H}$ is seperable in two matrices, one can write the latter formula as

$$\mathbf{w} = \left[(\mathbf{A}\mathbf{B})^{\mathbb{H}}\mathbf{A}\mathbf{B} + \gamma_r \mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\right]^{-1} (\mathbf{A}\mathbf{B})^{\mathbb{H}}\mathbf{p}. \tag{2.90}$$

2.4.3 Limit the Panning Function Order

The utilized microphone type determines the maximum mode order which can be depicted after recording. In practice, only finite order microphones are feasible, i.e.,

it is necessary to limit the bandwidth of the panning function. As the solution of the last section has no constraint on the bandwidth, it is desirable to limit the order, i.e., the bandwidth of the panning function. One can use a Fourier row expansion to depict the function with the help of sin and cos terms up to order M. The solution for the loudspeaker weightings for the source angles ϕ_s with $s = 1...S$ is then given by [Pol07]

$$\mathbf{W} = \mathbf{Q}\mathbf{T} \tag{2.91}$$

or more precisely as

$$\begin{pmatrix} w_1(\phi_1) & \cdots & w_1(\phi_S) \\ \vdots & \ddots & \vdots \\ w_L(\phi_1) & \cdots & w_L(\phi_S) \end{pmatrix} = \mathbf{Q} \begin{pmatrix} 1 & \cdots & 1 \\ \cos(\phi_1) & \cdots & \cos(\phi_S) \\ \sin(\phi_1) & \cdots & \sin(\phi_S) \\ \vdots & \ddots & \vdots \\ \cos(M\phi_1) & \cdots & \cos(M\phi_S) \\ \sin(M\phi_1) & \cdots & \sin(M\phi_S) \end{pmatrix}, \tag{2.92}$$

where $\mathbf{Q}$ is the $L \times 2M+1$ Fourier row coefficient matrix and $\mathbf{T}$ includes the sin and cos up to order M. To determine these coefficients, one can use the least squares solution [Pol07]

$$\mathbf{Q} = \mathbf{W}\mathbf{T}^{\mathbb{H}}(\mathbf{T}\mathbf{T}^{\mathbb{H}})^{-1}. \tag{2.93}$$

Chapter 3

Weighting Function Design

The Ambisonics theory provides defined rules for loudspeaker signal calculation as one can see from the Chap. 2. In the following section, the ITU loudspeaker constellation is analyzed and reasons are given why the theory does not give an optimal solution.

3.1 Weightings of the ITU Layout

The ITU (International Telecommunication Union) layout is a recommended loudspeaker arrangement for 5.1 sound systems. 5.1 systems typically consist of three front loudspeakers, left, right and center (L/C/R), and two surround speaker, left surround (LS) and right surround (RS) [IR12]. Furthermore, there is one low frequency effect channel without a defined position. As this layout is quite popular for the widely spread 5.1 system, it provides a good basis for comparable research regarding sound quality. Fig. 3.1 depicts the loudspeaker arrangement like it is defined in [IR12]. All five speakers possess the same radial distance to the origin. Three of them are in viewing direction of the listener, i.e., $\phi_l = \pm 30°$ and $\phi_l = 0°$. Two loudspeakers are positioned behind the listener at $\phi_l = \pm 110°$ to create the surround effect.

As shown in subsection 2.1.3, the number of loudspeakers limits the area and the frequency range in which the sound is reproduced properly. For irregular layouts, i.e., non-uniform distributed loudspeakers in a circular array, the correctly synthesized

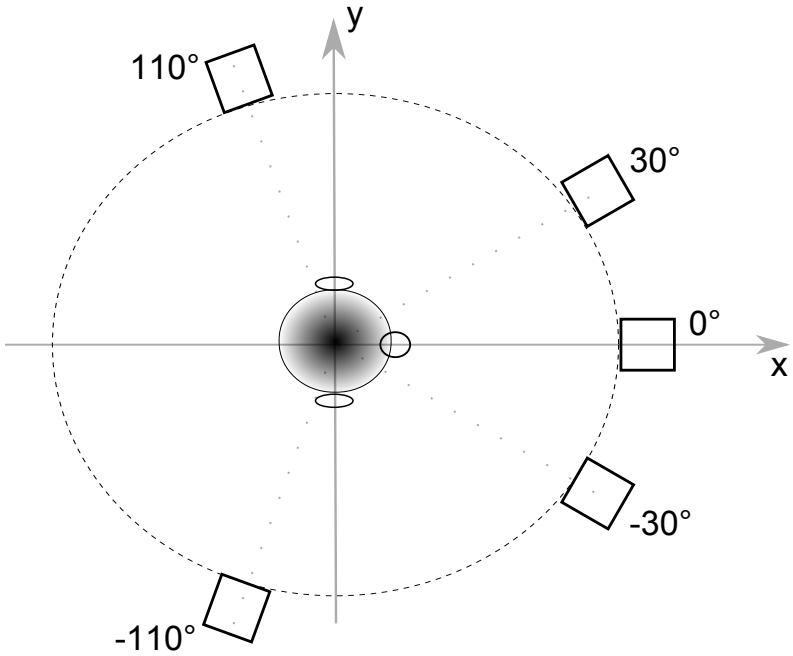

Figure 3.1: ITU reproduction layout as described in [IR12].

area is dependent on where the wave is coming from. The panning functions for irregular layouts are interpolants for periodic non-uniform sampling [Pol07]. A solution for non-uniformly sampled functions is derived in[ME04]. For the ITU layout, the theoretical solution for the signal weightings for each speaker are shown in Fig. 3.2. This figure depicts the amplitudes, the sum and the root mean square value at every angle $-180° \leq \phi_s \leq 180°$. Even if the sum is constant over the whole angular range, the amplitudes of the individual transducers reach partially enormous values. The theoretical computations are based on a correctly reproduced sound field in the origin. As long as the listener stays in the origin, he would perceive it correctly up to a certain frequency. As soon as he turns his head or moves just a bit, the enormous volumes of certain loudspeakers would distort the sound image extraordinarily. If one considers the angular range behind the listener one can see that all loudspeakers are active whereby three of them are driven with huge amplitudes. A movement of the listeners towards the frontal loudspeakers could create a frontal sound image perception. Furthermore, large loudspeaker weights mean that if there are slight differences between the loudspeaker

responses or variations of loudspeaker positioning, this will cause huge variations in the reproduced field. For that reason, the given solution for the ITU layout is non-robust. A useful method to increase the robustness would be to penalize loudspeakers in certain angular areas to limit the amplitude of these loudspeakers. The next section treats how to design a penalty function analytically.

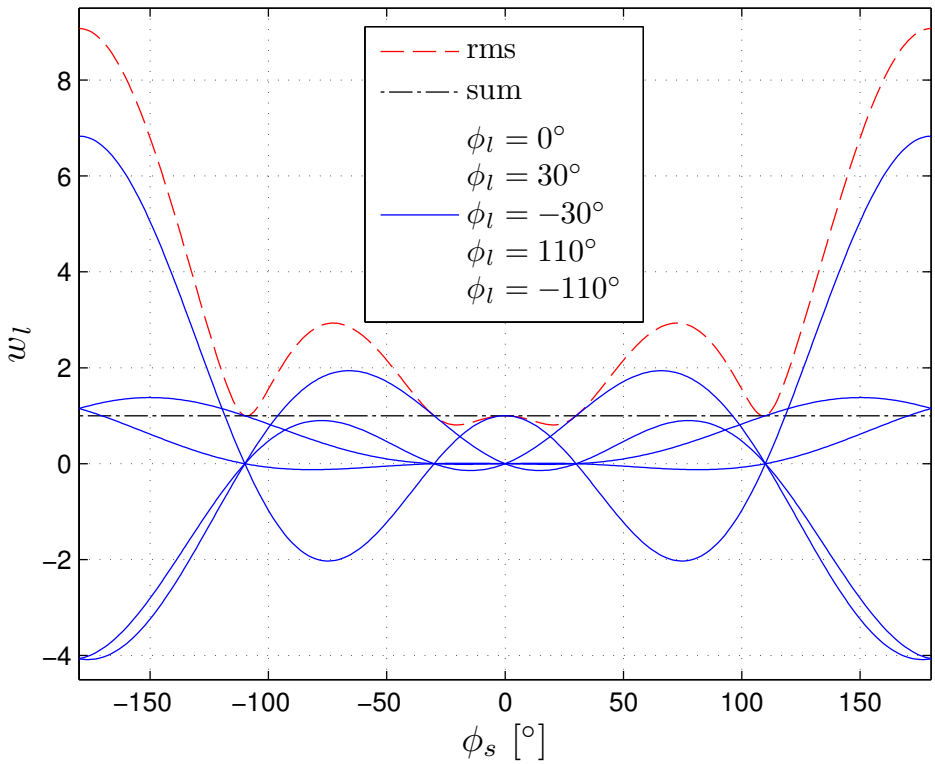

Figure 3.2: Interpolants for the ITU loudspeaker layout.

3.2 Penalty Function Design

3.2.1 Design

The penalty function can be seen as angular windowing of the loudspeaker. The angle of the source can be seen as the center of the window. Loudspeakers with small angular distance to the source play a bigger role for the reproduction of the desired wave. Before

designing a penalty function, the angular distance has to be defined as the difference between speaker and source angle

$$\Delta\phi = \phi_l - \phi_s, \tag{3.1}$$

where for $\Delta\phi = \pm 180°$ the desired and the loudspeaker wave would come from opposite directions, and for $\Delta\phi = 0°$ from the same direction. For the highest possible angular distance $\Delta\phi = \pm 180°$, it is desired to create a high penalty. For low angular distances where $\Delta\phi$ is close to zero, a penalty close to zero would be useful. A penalty which was already published in [Pol07] is defined as

$$\gamma(\Delta\phi) = 0.5(1 - \cos(\Delta\phi)). \tag{3.2}$$

The function is periodic, reaches one for loudspeakers from the opposite direction and is zero for $\Delta\phi = \pm 0°$. One disadvantage of the function is that it is not really controllable between 0 and 180 degree. A more sophisticated method would have several parameters to control the curve in this angle range. One attempt with similar properties regarding periodicity provides the function

$$\gamma(\Delta\phi) = 1 - e^{-b(1-\cos(\Delta\phi))^p}. \tag{3.3}$$

It has the same characteristics for input angles of 0 and 180 degrees. In addition, it provides several parameters to control the envelope in between. So what would be useful requirements for the curve shape? First of all, it is desirable to penalize the loudspeakers whose angular distance is more than $\Delta\phi = \pi/2$ with a value close to one. To ensure this, one can use the parameter b to adjust the behavior of the function so that the condition $\gamma(\Delta\phi = \pi/2) = 1 - \epsilon$ is fulfilled, where ϵ should be a value close to zero. Equaling

$$1 - \epsilon = 1 - e^{-b(1-\cos(\pi/2))^p} \tag{3.4}$$

and resolving for b, one receives

$$b = -\ln \epsilon. \tag{3.5}$$

For $\epsilon = 0.01$, which is close to zero, b has to be set to about 4.6. For the rest of this thesis, it is set to $b = 4$ which results in $\epsilon = 0.0183$. The value p in Eq. 3.3 controls the steepness of the curve, i.e., how fast the value defined by ϵ is reached. In Fig. 3.3, the penalty function curves for Eq. 3.2 and Eq. 3.3 are depicted, where $p \in \{1, 2, 4\}$ are illustrated.

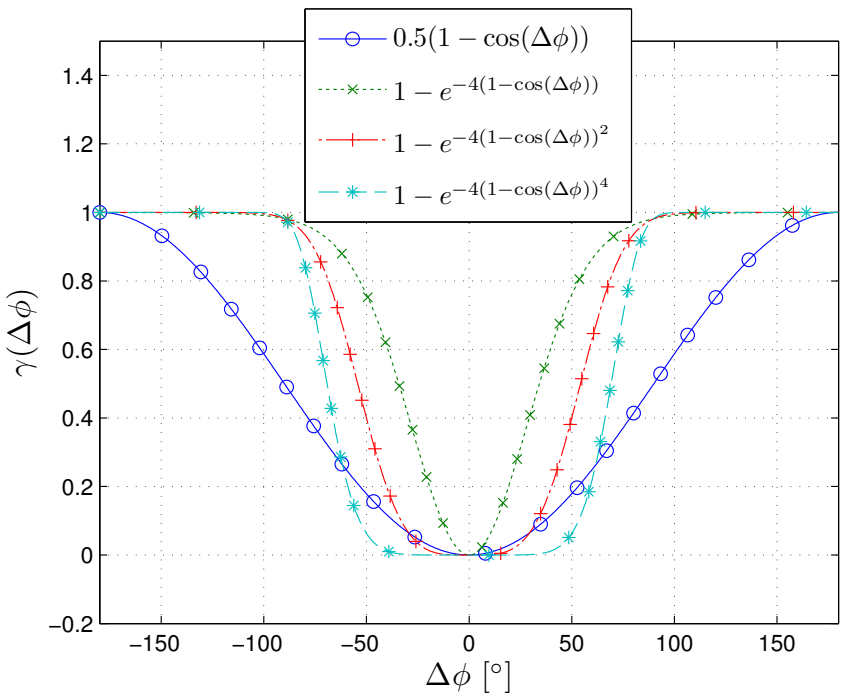

Figure 3.3: Plot of several penalty functions with low angular distance in the center and high angular distance on the left/right side.

Another way to assign a penalty would be to give no penalty to the closest loudspeakers in terms of angular distance. But in some cases, this condition does not fulfill an intuitive selection for the penalty assignment. If one assumes three loudspeakers with the same radius and with angle towards the origin 0°, 10° and 30°, and a desired sound source coming from 14°, the two nearest loudspeakers would be at 0° and 10°. In this case the intuitive loudspeaker selection would be 10° and 30° to synthesize the desired

wave, i.e., to give the penalty to loudspeaker at 0°. Intuitively, it would be more useful to give no penalty to one loudspeaker in positive angular distance and to one in negative angular distance, where the angular distance is defined as $-\pi < \Delta\phi \leq \pi$. The mathematical formulation of this rule is defined as

$$\gamma(\Delta\phi) = \begin{cases} 0 & \text{if } \Delta\phi = \min\limits_{\phi_l}(\phi_l - \phi_s) \text{ for } \Delta\phi \geq 0 \\ 0 & \text{if } \Delta\phi = \min\limits_{\phi_l} |\phi_l - \phi_s| \text{ for } \Delta\phi \leq 0 \\ 1 & \text{else .} \end{cases} \tag{3.6}$$

The idea behind this penalty function is to attenuate all loudspeakers except the two where the desired sound source is in between. This is the basic idea of VBAP used e.g. in [Pul97].

3.2.2 Penalty Function Analysis

Cosine Penalty

In the last section, several possibilities for penalty functions have been given. These functions provide a good basis for further analysis, which is given in this section with reference to the ITU layout introduced in Sect. 3.1. Fig. 3.4 depicts the driving (panning) functions for all of the five loudspeakers. The left plot shows the unlimited solution of Eq. 2.89. The right plot shows the driving functions limited up to order M as defined by Eq. 2.93. It is obvious that the amplitudes stay below 1.3 and over -0.3 and do not reach enormous values as the ITU interpolants do, which are shown in Fig. 3.2. This is based on the fact that the total power is regulated. Furthermore the side lobes are reduced because of the angular penalty. The sum of the amplitudes stays around one for all angles and especially the root mean squared (rms) value, which is an indicator for the used power, remains close to one in marked contrast to the rms power in Fig. 3.2. The found driving functions create a much more robust sound field. In [Pol07], further analysis regarding the synthesized sound field is given. For example, the radial/frequential error increases is much lower with the penalty condition. In the

scope of this thesis, the resulting sound field is evaluated with the help of a listening test which is described in 5.

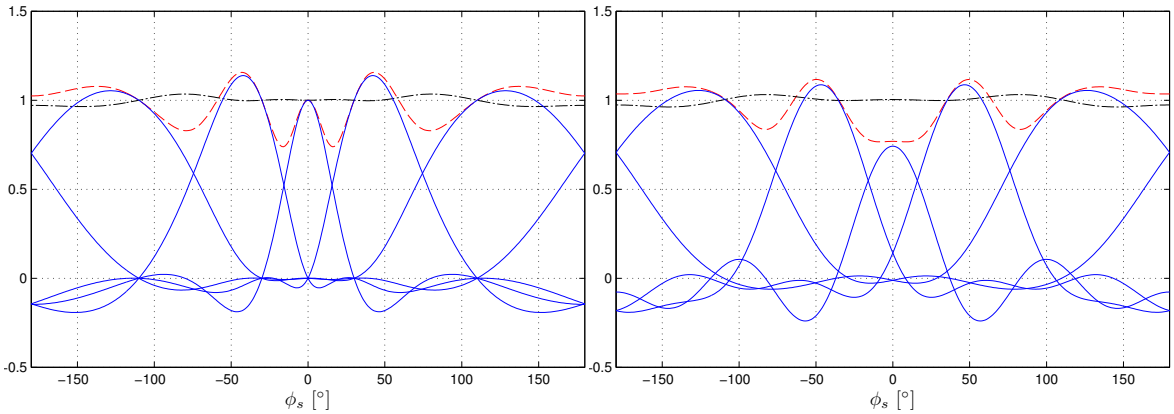

Figure 3.4: Panning functions for the ITU layout regularized by $\gamma_r = 1.5$ with the cosine penalty function. Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights.

Exponential Penalty

Besides the cosine penalty, a function with exponential behavior was introduced in the last section in Eq. 3.3. Fig. 3.5 illustrates the plots for the ITU layout driving functions created with the regulated least squares solution using the exponential penalty function. On the left side one can see again the unlimited solution and on the right side the order limited one. Compared to the cosine penalty, the side lobes in the frontal speaker area are better attenuated for the left and the right curve. The area behind the listener develops a similar shape. Summarizing one can say, this setup would also be suitable for a listening test but has not been considered.

As the exponential penalty function provides several parameters to control its influence on the driving function, another parameter set is shown in Fig. 3.6. The arrangement of the plots is the same as before. In this case, the high exponent causes noticeable side lobes. Furthermore, the shape of the curve seems to be more abrupt caused by the nonlinear influence of the exponential penalty for the unlimited solution. This behavior disappears as one would expect for the order limited version. The amplitudes

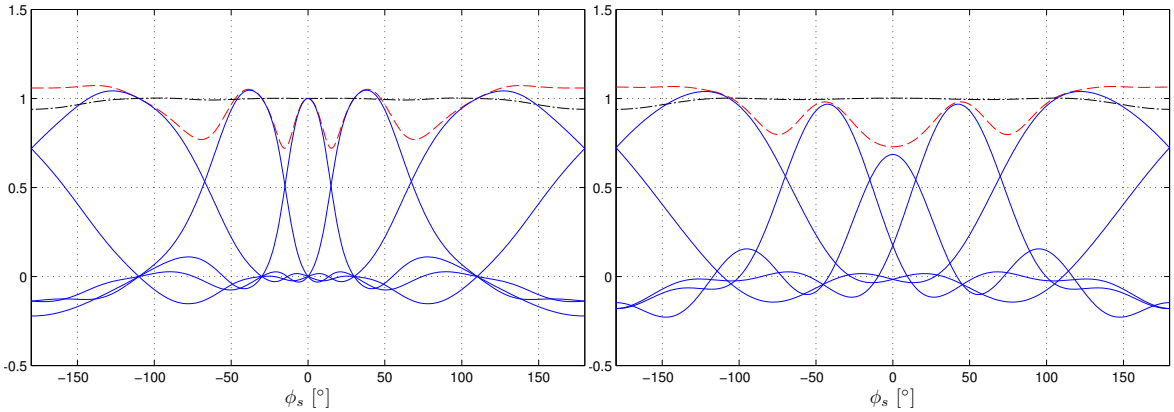

Figure 3.5: Panning functions for the ITU layout regularized by $\gamma_r = 0.15$ with the exponential penalty function ($b = 4, p = 1$). Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights.

have some obvious deviations from one that would create a non robust sound image. For that reason, it would not be suitable for listening tests.

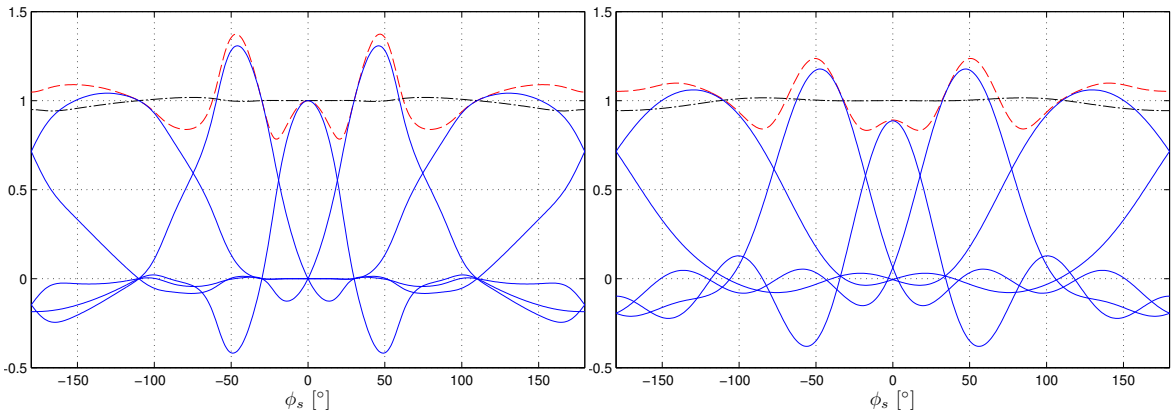

Figure 3.6: Panning functions for the ITU layout regularized by $\gamma_r = 0.5$ with the exponential penalty function ($b = 4, p = 4$). Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights.

Pairwise Penalty

The last penalty constraint which has been introduced is the pairwise one defined in Eq. 3.6. Fig. 3.7 depicts the panning functions for every of the five loudspeakers for both solutions. In contrast to the plots before, the shape of the curves is clearly non-smooth. For this case, the penalty constraint lowers the side lobes and attenuates

the non preferred loudspeaker amplitudes. For higher values of the regularization parameter γ_r at most two speakers are active for reproduction of a wave coming from a desired direction. This kind of reproduction method is used for the listening test described later. In this case, the order limited solution does definitely not provide the desired result. In the next section, the connection of the pairwise penalty to the traditional stereo panning laws are given.

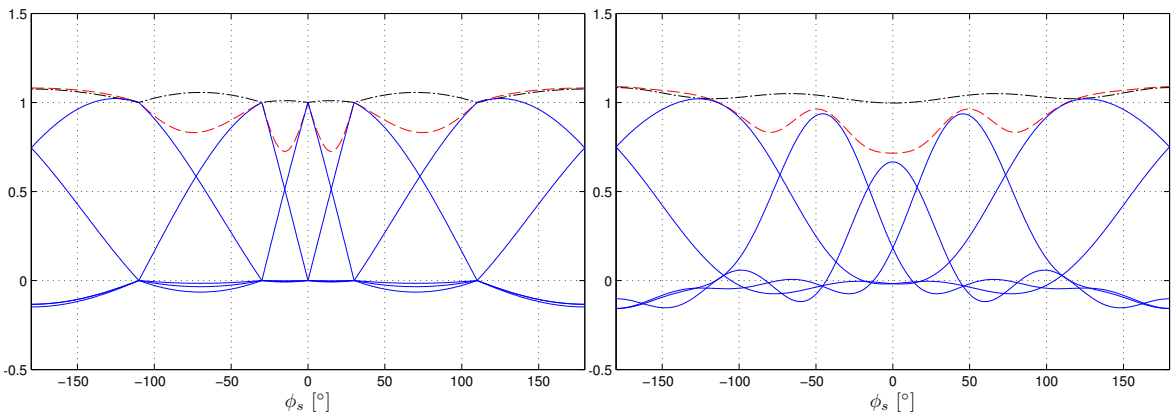

Figure 3.7: Panning functions for the ITU layout regularized by $\gamma_r = 2$ with the pairwise penalty function. Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights.

3.2.3 From Pairwise Penalty to the Stereo Panning Laws

As shown in Fig. 3.7, for high values of regularization factor γ_r, there are at most two loudspeakers active during the replay. The question is which stereo panning law is used to create the image of the virtual source, i.e., how the presented solution connected is to the classical stereophony [Sno55]. By showing the influence of γ_r on the loudspeaker solution Eq. 2.89, the weightings for the two active loudspeaker can be derived. First of all, one defines separately the terms of

$$\left[\mathbf{H}^{\mathbb{H}}\mathbf{H} + \gamma_r \mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\right]^{-1}, \tag{3.7}$$

that is

$$\mathbf{H}^{\mathbb{H}}\mathbf{H} = \begin{pmatrix} 1+2\cos(\phi_1-\phi_1) & 1+2\cos(\phi_2-\phi_1) & \cdots & 1+2\cos(\phi_L-\phi_1) \\ 1+2\cos(\phi_1-\phi_2) & 1+2\cos(\phi_2-\phi_2) & \cdots & 1+2\cos(\phi_L-\phi_2) \\ \vdots & \ddots & \ddots & \vdots \\ 1+2\cos(\phi_1-\phi_L) & 1+2\cos(\phi_2-\phi_L) & \cdots & 1+2\cos(\phi_L-\phi_L) \end{pmatrix}, \quad (3.8)$$

and

$$\mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \quad (3.9)$$

For large values of γ_r, the penalized mode Eq. 3.7 can be expressed in form of

$$\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{bmatrix}^{-1}, \quad (3.10)$$

where $\mathbf{D}$ can be approximated with a diagonal matrix with large values on the diagonal. Considering this blockwise writing, the solution for the latter equation is given as in [Hag89]

$$\begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{U}\mathbf{C}^{-1}\mathbf{V}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{U}\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\mathbf{V}\mathbf{A}^{-1} & \mathbf{C}^{-1} \end{bmatrix}, \quad (3.11)$$

with $\mathbf{C} = \mathbf{D} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}$. The inverse of $\mathbf{C}$ can be approximated with the inverse of $\mathbf{D}$ because of its large values on the diagonal. The inverse is then a zero matrix $\mathbf{0}$. Keeping this approximations in mind, one receives

$$\left[\mathbf{H}^{\mathbb{H}}\mathbf{H} + \gamma_r\mathbf{\Gamma}^{\mathbb{H}}\mathbf{\Gamma}\right]^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (3.12)$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{H}^{\mathbb{H}}\mathbf{H}$ for the first two loudspeakers, whatever their angles $\phi_{1/2}$ are. As one can see, the mode matching is in this case limited to the two loudspeakers, where the source is in between. The matching equation in this case is defined as

$$\begin{pmatrix} e^{\mathrm{i}\phi_s} \\ 1 \\ e^{-\mathrm{i}\phi_s} \end{pmatrix} = \begin{pmatrix} e^{\mathrm{i}\phi_l} & e^{-\mathrm{i}\phi_l} \\ 1 & 1 \\ e^{-\mathrm{i}\phi_l} & e^{\mathrm{i}\phi_l} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{3.13}$$

By subtracting and adding the first and the third line, this matching equation can be written as

$$\begin{pmatrix} 1 \\ \cos\phi_s \\ \sin\phi_s \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \cos\phi_l & \cos\phi_l \\ \sin\phi_l & -\sin\phi_l \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{3.14}$$

The latter system of equation can be rewritten as

$$\begin{pmatrix} I \\ II \\ III \end{pmatrix} \begin{matrix} 1 \\ \cos\phi_s \\ \sin\phi_s \end{matrix} \begin{matrix} = \\ \\ \\ \end{matrix} \begin{matrix} w_1 + w_2 \\ (w_1 + w_2)\cos\phi_s \\ (w_1 - w_2)\sin\phi_s \end{matrix} \tag{3.15}$$

Equation III divided by II results in the tangent law for stereophony [RS09]

$$\tan\phi_s = \frac{(w_1 - w_2)}{(w_1 + w_2)} \tan\phi_l, \tag{3.16}$$

which has been proposed earlier in [Ber73]. Another well known stereophonic law is the law of sines which was derived in [Bau62] and can be achieved by dividing equation III and I

$$\sin\phi_s = \frac{(w_1 - w_2)}{(w_1 + w_2)} \sin\phi_l. \tag{3.17}$$

Fig. 3.8 shows a comparative plot for the sine, tangent and mode matching panning stereo law for the two loudspeakers at $\phi_l = \pm 30°$. As one can see, there are just minor differences in the driving functions. The mode matching condition creates a panning law which lies exactly between sine and tangent.

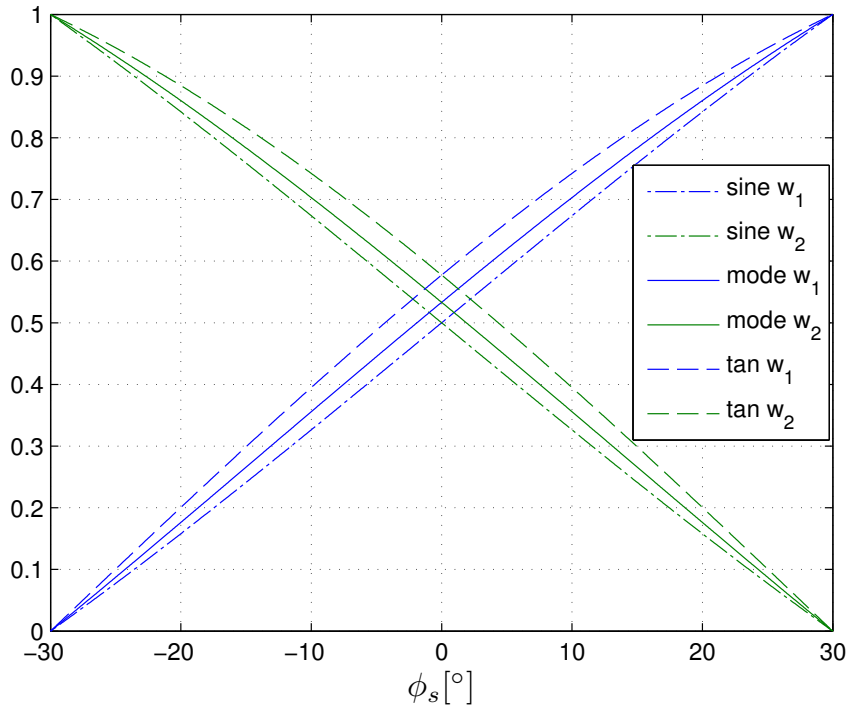

Figure 3.8: Sine-, mode and tangent panning law in comparison for loudspeakers at $\phi_l = \pm 30°$.

3.3 Frequency Dependent Design

The idea of creating a frequency dependent weighting to optimize the spatial sound image was already mentioned by Gerzon in [Ger85], where he uses a shelf filter to match the different requirements given by the spatial perception of humans for the frequencies below and above 700Hz. As Gerzon says, the human head attenuates acoustical waves for frequencies above 700Hz. Another frequency dependent approach was done by Daniel in [Dan03], where near-field effects of the reproduction loudspeakers are compensated using distant-coding filters. Within the scope of this thesis, the frequency-

dependent design is based on the limitation given by the number of loudspeakers. The next section shows how the reproduction setup influences the frequency-dependent weighting function design.

3.3.1 Robust Panning for given Reproduction Setups

As mentioned previously, the limited number of loudspeakers results in restrictions for sound field synthesis. These restrictions result in the order limitation for the mode matching which leads to bounds for frequency and radius in which the sound field is reproduced properly. For the ITU layout, the number of loudspeakers is $L = 5$, the mode matching order is set to $M = 1$ using Eq. 2.19 or $M = 2$ using Eq. 2.18. Assuming the minimum requirement is the correct reproduction of the sound field around the human head, which has a radius of about $R_h = 8.75$cm [BD98] [Kuh77], one receives for the upper frequency limit by using Eq. 2.15 $f_{max} \approx 1248$Hz for an order $N = 2$ or $f_{max} \approx 624$Hz for an order $N = 1$. The maximum frequency represents the theoretical bound in which the sound field is synthesized correctly within the head radius around the origin for given error conditions of Eq. 2.18 or Eq. 2.19. A frequency dependent weighting function design can be developed based on the cues provided by the Ambisonics theory. Below that frequency, this theory holds and is applicable, above only pairwise loudspeakers should be used for the reproduction which leads to stereo panning laws. Having defined the cue points for the frequency ranges, one can define the requirements for the filters.

3.3.2 Interpolation between the required Weighting Factors

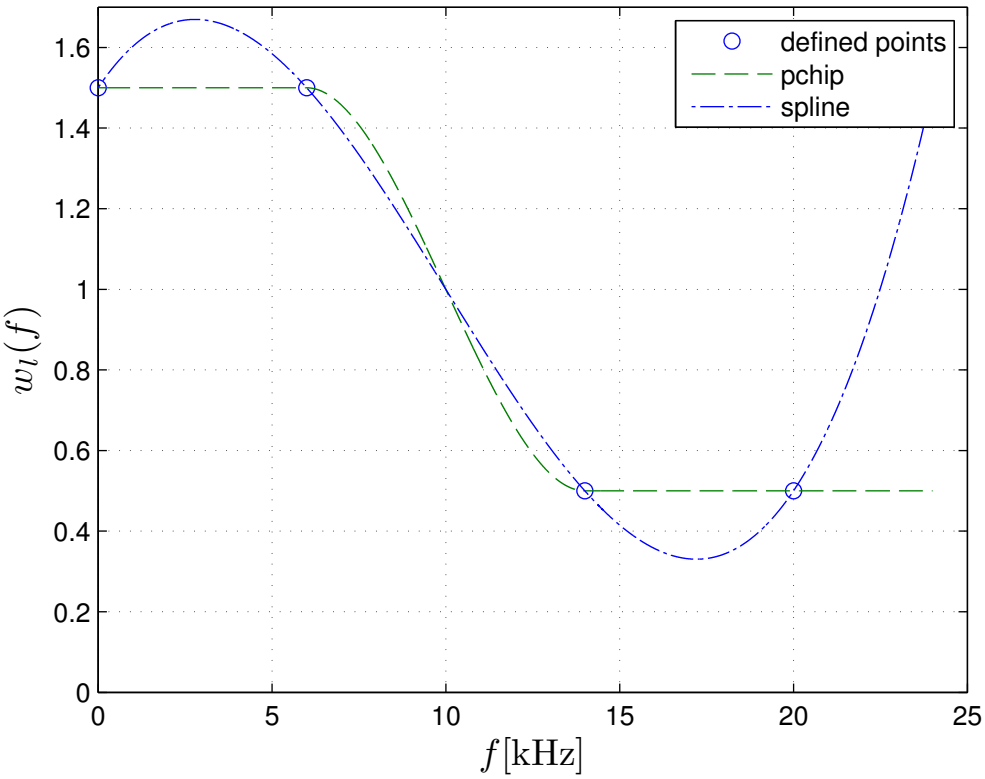

Figure 3.9: Plot of two different interpolation methods for four defined points.

One wants to derive panning functions that provide the largest possible reproduction area at low frequencies, but it is required to limit the effects of side-lobes at high frequencies, where the reproduction radius is smaller than the human head. Therefore it is desirable to change the panning functions with frequency. The frequency-dependent cue points have been developed using the Ambisonics theory. These points define weighting factors for the loudspeaker signals on the frequency axis. Analytically, one can define two required levels, one given by the Ambisonics theory below the boundary frequency and one level calculated by stereo panning laws for the upper frequencies. This would result in something like low or high pass at a cutoff frequency with defined pass, transition, and stop band. In the scope of this thesis, the main issue was the design of the frequency response at given points. The right choice of the filter would need

some more research. To create a smooth transition between the determined points, one can use simple interpolation functions, which are for example provided by Matlab. Fig. 3.9 shows the behavior of two different interpolation methods provided by the Matlab library. Beneath the two shown methods, the linear interpolation has been shortlisted for the frequency response creation. The linear function has non-smooth behavior on the change to the transition band which reduces the decay of the impulse response. The spline interpolation causes unwanted overshooting in the pass and stop band. An optimal behavior provides the 'pchip' (piecewise cubic Hermite interpolating polynomial) [Mol13] spline which uses a piecewise cubic Hermite interpolation. The amplitude stays constant in the pass band between the selected points and does not cause distortions. Additionally, the transition to the second amplitude behaves smoothly. A mathematical description can be found in [Mol13] and implementations in [KMNF89].

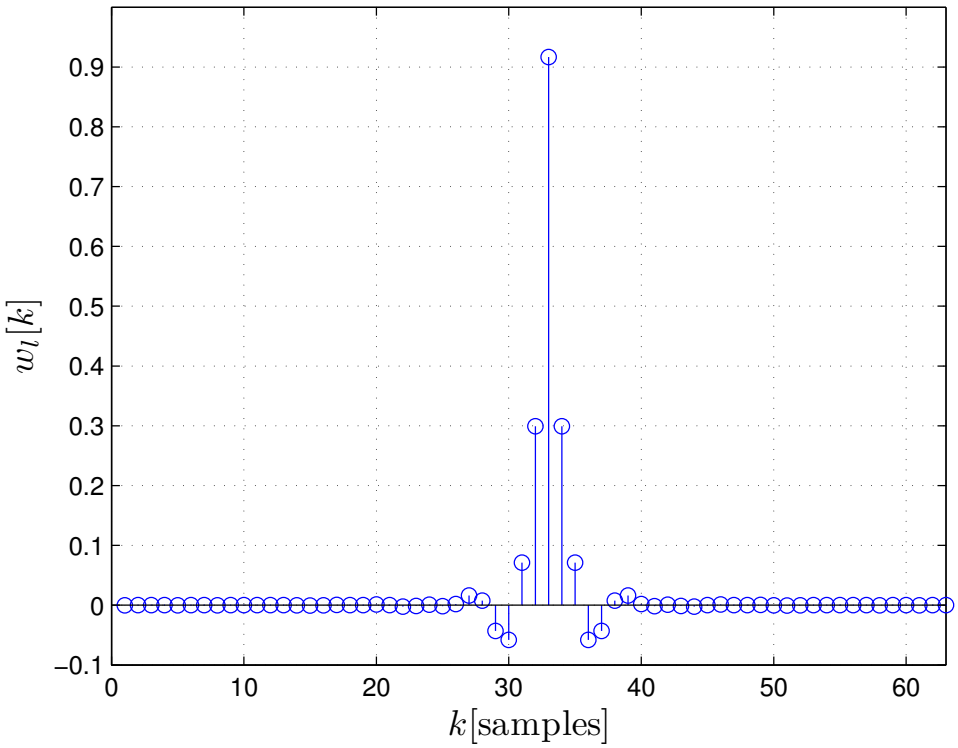

Figure 3.10: Exemplary plot of an impulse response with 64 samples.

After using an inverse fast Fourier transformation (IFFT), one receives the real impulse

response (IR) of the linear phase filter, which is plotted in Fig. 3.10. It is desirable to have a real impulse response as the panning functions are real. This was achieved by complex conjugated mirroring of the frequency response at $f_s/2$. It is obvious that most of the values are zero. A more efficient way to represent this IR is shown in the next section.

3.3.3 Reducing the Impulse Response Coefficients

After defining a decent frequency envelope for the weightings, it is desirable to reduce the number of coefficients of the impulse response, which results in a smaller group delay, less data and a smaller computational effort. The discrete impulse response samples are connected to the z-domain via the z-transform, which is the counterpart to the Laplace transform for discrete values and is defined as [OSB99]

$$G(z) = \sum_{k=-\infty}^{\infty} g[k]z^{-k}. \tag{3.18}$$

For a simple first-order linear phase filter with a symmetric impulse response

$$g[k] = a_1\delta[k] + a_0\delta[k-1] + a_1\delta[k-2], \tag{3.19}$$

one receives the z-transform

$$G(z) = a_1 z^0 + a_0 z^{-1} + a_1 z^{-2}. \tag{3.20}$$

For

$$z = e^{\mathrm{i}\Omega}, \tag{3.21}$$

$G(z)$ can be expressed as

$$G(e^{\mathrm{i}\Omega}) = a_1 e^{\mathrm{i}\Omega 0} + a_0 e^{-\mathrm{i}\Omega} + a_1 e^{-\mathrm{i}2\Omega}, \tag{3.22}$$

where $\Omega = 2\pi f/f_s$ is the normalized angular frequency with the sampling frequency f_s. $G(e^{\mathrm{i}\Omega})$ is also called discrete time Fourier transform (DTFT) [OSB99]. The latter equation can now be expressed with cosine terms as

$$G(e^{\mathrm{i}\Omega}) = (a_0 + 2a_1 \cos\Omega)e^{-\mathrm{i}\Omega}. \tag{3.23}$$

The magnitude of the frequency response and the weightings can therefore be expressed with cosine terms. This is again a Fourier series expansion, where it is sufficient to calculate the coefficients for cosine terms, because the frequency response should be an even function around $\Omega = 0$. The weightings $\mathbf{W}$ at the frequencies Ω_μ with $\mu = 1...K$ which are created by cosine terms up to order M can be expressed as

$$\mathbf{W} = \mathbf{QT}, \tag{3.24}$$

or more precise as

$$\begin{pmatrix} w_1(\Omega_1) & \cdots & w_1(\Omega_K) \\ \vdots & \ddots & \vdots \\ w_L(\Omega_1) & \cdots & w_L(\Omega_K) \end{pmatrix} = \mathbf{Q} \begin{pmatrix} 1 & \cdots & 1 \\ \cos(\Omega_1) & \cdots & \cos(\Omega_K) \\ \vdots & \ddots & \vdots \\ \cos(M\Omega_1) & \cdots & \cos(M\Omega_K) \end{pmatrix}. \tag{3.25}$$

The least squares solution for $K > M + 1$ is then given by

$$\mathbf{Q} = \mathbf{WT}^{\mathbb{H}}(\mathbf{TT}^{\mathbb{H}})^{-1} = \begin{pmatrix} a_0^1 & \cdots & 2a_M^1 \\ \vdots & \ddots & \vdots \\ a_0^L & \cdots & 2a_M^L \end{pmatrix}, \tag{3.26}$$

where a_M^L is the Mth coefficient of the Lth loudspeaker.

This section has shown how to generate a simple linear phase filter from a predefined frequency response defined at crucial points and then interpolated. In the next chapter, a tool is introduced which allows to define these points and to perform an audio demonstration afterwards.

Chapter 4

Audio Control Interface

Within the scope of this thesis, an Ambisonics decoder has been developed for sound field synthesis based on a free field audio scene. To control the behavior of the decoder, a graphical user interface (GUI) has been created. The next section explains adjustable features of the GUI.

4.1 Graphical User Interface

The filter coefficients have been defined with the aid of crucial points in the frequency domain. At first, it would be interesting which parameters are necessary to determine one of these points. Fig. 4.1 shows the interface in terms of an editable table which has been created with the help of Matlab. In the first row one can see an editable text field to name the point. The second row marked with 'Penalty Function' offers a drop down menu with predefined penalty functions which have been introduced in Sect. 3.2. For users who want to create their own penalty function, the next row provides an editable text field in which one can type in an individual penalty function in Matlab syntax provided the check box left of it marked with 'Design Penalty' is activated. The term 'Regularization' refers to the factor γ_r controlling the influence of the penalty function on the loudspeaker solution Eq. 2.89. It can be adjusted with the editable text field. In the fifth row, one is able to limit the order of the trigonometric functions, the weighting

functions are represented with. If the value 'Weighting Function Order' is zero, there is no limitation on the order. The matching order can be controlled with the text field in the following row. Useful results can be achieved with the coherence of Eq. 2.18 or Eq. 2.19 whereby the latter represents the lower limit. A very important parameter is the frequency at which the interpolation is created. For a set of interpolation points, there is only one point possible at a single frequency. The last row allows with the help of a drop-down menu the insertion of external weightings like the ones of Peter Craven which he has created on the basis of various psychoacoustic criteria [Cra03]. The following listing shall summarize the just described parameters:

- 'Name': Name of the interpolation point.
- 'Penalty': Drop-down menu with several penalty function suggestions.
- 'Design Penalty': If the check-box is activated, one can create an own function in the text-field using Matlab syntax.
- 'Regularization': This value corresponds to the factor γ_r in Eq. 2.89 which defines the influence of the penalty on the solution.
- 'Weighting Function Order': Maximum order of the trigonometric functions which represent the weightings.
- 'Mode Matching Order': Maximum order of the modes which has to be matched to find the solution as defined in Eq. 2.18 or Eq. 2.19.
- 'Frequency': The frequency at which the interpolation point is created.
- 'External Weightings': If a check-box is activated, an external weighting is used for the interpolation point.

If the user changes the parameters in the GUI, the plot for the weighting functions is updated simultaneously. If one clicks on the button 'Create Interpolation Point', the

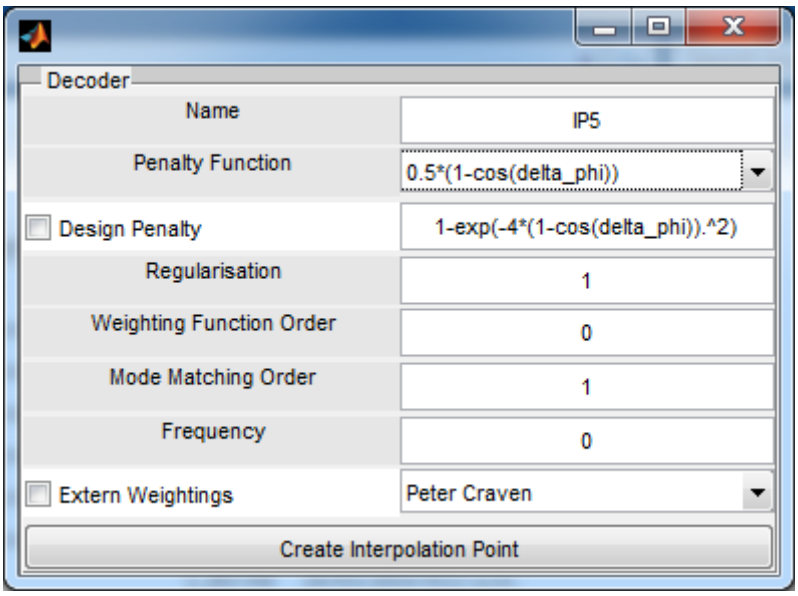

Figure 4.1: Window for the creation of an interpolation point.

interpolation point is calculated for the given parameters at the adjusted frequency and saved for one decoder.

Beneath the weighting function plots, Fig. 4.2 shows two decoders and their created interpolation points. Every decoder possesses buttons for creating, editing and deleting the point selected in the list, and for saving or loading a set of points. Clicking on a point in the list of one of the decoders will plot the corresponding weighting functions on the right side. For the upper decoder, four interpolation points have been created. As shown in Sect. 3.3, two points define the lower frequency range and two points the upper frequency range. For the upper frequency range, only pairwise panning provides robust results that is at most two loudspeakers are active which can be seen in the upper plot. In the central upper part of Fig. 4.2, one can see the loudspeaker positioning and a slider to control the total volume of the rendition. The positioning of the reproduction setup is controlled with a file in extensible markup language (XML) syntax. Another feature regarding the playback of the audio demonstration is displayed in the lower left of the window. Beneath a start button, a button pause and stop reproduction is provided. The other elements in the lower left are:

- 'Angle Range': Defines the angular range in which the audio demo is performed.
- 'Angle Steps': Defines the step size within the angular range for the reproduction of the test signal.
- 'Signal Time': Duration of the signal at the defined angles.
- 'Signal Type': Drop-down menu with several test signal types.
- 'Pause after signal': Pause after rendition of the test signal with Decoder 1/2.
- 'Edit Signal Filter': Apply a filter on the defined signal.

The next section dwells on how the test signal is created and what kind of test should be used for listening tests which aim on spatial perception.

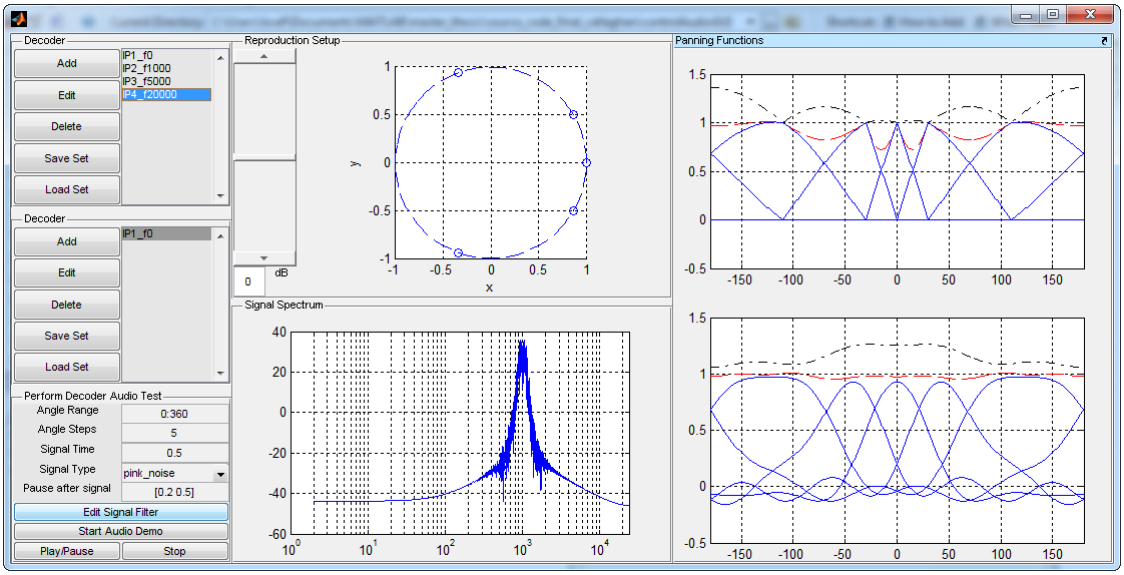

Figure 4.2: Interface for the control of the decoder and the listening test.

4.2 Test Signal Creation

Besides usability regarding the decoder parameter and the audio demo, it would be useful to have a variety of test signals. Different kind of signals have been used to

analyze the localization. Typically, 1/3 octave band noise [Bla97] or pink noise [PH05] lead to useful spatial perception and create enough cues for the human hearing. The requirements for these frequency bands are defined in standard ANSI SI.II-1986 [TB86]. The GUI allows to select either a sine, pink or white noise and to read a whole WAVE audio file with the help of a drop-down menu. After defining the signal, it is possible to apply a filter via the 'Edit Signal Filter' button. After clicking on the button, a window like in Fig. 4.3 opens and one is able to choose between low, band (1/3 octave or octave) or high pass with a given mid/cut-off frequency. To meet the requirements of standard ANSI SI.II-1986 [TB86], a Butterworth filter of order four has been used. In Fig. 4.2, one can see the Fourier transform of pink noise filtered with 1/3 octave band Butterworth filter.

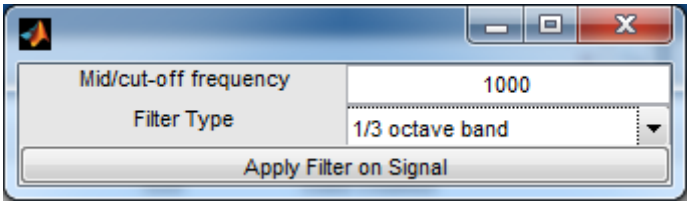

Figure 4.3: Interface for applying a filter on the created signal.

Having introduced a useful tool to perform a listening test, the next section talks about how the test is prepared and conducted.

Chapter 5

Listening Test

Typically, listening tests should be conducted under free field conditions that is in an anechoic environment. In the scope of this thesis, it has not been possible to provide such conditions. Nevertheless, the first incoming wave dominates the perception, i.e., sound incoming up to 50ms after the first wave does not have an influence on the localization. This fact is also known as precedence effect [LCYG99] and was already mentioned in 1951 [Haa51]. Besides the spatial perception, the room size has an influence on the power of the reflected sound. The following section deals with the room geometry and its influence on the reflections. Furthermore, the positioning of the reproduction setup is described.

5.1 Room Description and Experimental Setup

The listening test was not performed in an anechoic chamber, even if the theoretical derivation of the driving functions assumes free-field conditions. In the following, a short description of the room shape and the resulting assumptions for the reverberant field are given. The reverberation time, which is the time an acoustical impulse requires to loose 60dB sound pressure level, was measured $T_{60} = 0.7s$. The room has height has a height of 3.3m, a width of 4.3m and a length of 6.1m, and so the volume is approximately $V \approx 3.3 \cdot 4.3 \cdot 6.1\text{m}^3 \approx 86.6\text{m}^3$. Fig. 5.1 shows a picture of the room where

the listening test was conducted. With a given room geometry and a reverberation time, the acoustic absorption A_a measured in Sabin can be calculated as [KFCS99] [p. 336]

$$A_a = 0.161 \frac{V}{T_{60}} \text{ [Sabin]} . \tag{5.1}$$

Figure 5.1: Picture of the room with reproduction setup and a dummy head on the position of the listener.

This absorption factor is necessary to compute the reverberant intensity I_r of a sound field [KFCS99] [Ber54]

$$I_r = \frac{4P_a}{A_a c}, \tag{5.2}$$

where P_a is the acoustic power output of a sound source. The direct intensity of a sound source with directivity factor D is given as [KFCS99] [PFN10] [AS00]

$$I_d = \frac{P_a}{4\pi c r^2} D. \tag{5.3}$$

To calculate the radius around a sound source, in which the direct intensity is higher than the reverberant, one equals the two latter terms

$$\frac{P_a}{4\pi c r^2} D = \frac{4P_a}{A_a c}. \tag{5.4}$$

Resolving for the radius

$$r^2 = \frac{A_a}{16\pi} D \tag{5.5}$$

and using the definition of the acoustic absorption, one receives

$$r \approx 0.057 \sqrt{\frac{V}{T_{60}} D}\, [m]\,. \tag{5.6}$$

As expected, it is desirable that the position of the listener is within this radius for every loudspeaker of the reproduction setup. In this case, the radial position is the same for every loudspeaker as defined by the ITU layout. The directivity factor for a piston source (loudspeaker sound production can be approximated with the movement of a piston) is defined as [KFCS99]

$$D = \frac{(ka)^2}{1 - J_1(2ka)/ka}, \tag{5.7}$$

where a is the radius of the piston, k is the wave number and $J_1()$ is the Bessel function [Ber54] of the first kind of first order. For a piston radius $a = 0.09$m and loudspeaker radius of about 1 meter, the reverberant is equal to the direct field for a frequency of around 600Hz. Considering this at the listener position, the direct field has the bigger influence on the sound field perception of the listener for frequencies over 600Hz. The shape of the room and the resulting sound reflections are somehow determined and unchangeable. However, it is possible to avoid distortions of the synthesized sound field with accurate positioning of the loudspeakers. The radius for all loudspeakers has been chosen to one meter to keep some distance to the listener and to the walls. A problem with angular positioning arises from a certain loudspeaker size. Furthermore, it is easier to measure straight lines with the help of a tape measure than measuring angles, especially for circles with a radius of one meter. For that reason the center loudspeaker of the ITU layout has been positioned first with the given radius. Afterwards, all other

loudspeakers have been positioned on the same radial distance with the help of the chord, which transforms an angular distance to a direct distance and is defined as

$$d_c = 2r \sin\left(\frac{\alpha}{2}\right), \tag{5.8}$$

whereby α is the angular distance between the two loudspeakers. For the reproduction, the high resolution studio monitor 'hr824' from the the company MACKIE has been used [1]. The following section deals with the implementation and test items which have been played at the listening test.

5.2 Implementation and Test Items

The methods for deriving solutions for the loudspeaker signals mentioned in the previous chapters provide a variety of weighting functions. Besides the introduced methods, already published methods [Cra03] have been used. The test has been performed as a one to one comparison that is two items have been played in row at the same angle and after a short pause played at the next angle. In total, it has been four item pairs in four angular ranges for five participants. The listener was asked to stay in the center of the loudspeaker array sitting on a chair with view towards the center loudspeaker of the ITU layout. After playing an item pair for one angular range, the listener should describe the items in comparison to each other regarding smooth movement when changing the angle, a stuck image towards a loudspeaker, signal distortion, wideness of the image or a perceived front/back confusion. The listeners who took part have been working in the audio research area for some time. Unfortunately, the number of participants was too small to achieve statistical significance, yet the results are useful in indicating how the used panning functions perform. The results shall just give an idea how the methods behave properly regarding the reproduction and give reason for a more sophisticated listening test. Since the most advanced method presented in this thesis is the frequency dependent weighting, it is interesting to know

[1] `http://www.mackie.com/products/hr824/specs.html`

how it affects the spatial perception. For that reason it is included as the main item of the listening test. It consists of a cosine penalized weighting for frequencies up to 1000 Hz and a pure pairwise amplitude panning for the upper frequencies. Additionally, both weighting methods have been included as items separately to compare it with the combined solution. To classify the elaborated techniques with already published weightings for the ITU layout, the panning functions of Peter Craven [Cra03] complete the list:

1. Frequency dependent weighting function with angular penalty like the example described in Sect. 3.3: PenF.

2. Peter Cravens panning functions for the ITU layout [Cra03]: Craven.

3. The weighting functions created by the pairwise penalty as described in Sect. 3.2: PW.

4. The weighting functions created by the angular penalty as described in Sect. 3.2 or in [Pol07]: Pen.

The abbreviations shall simplify further identification of the items. As a first comparison, the panning functions of Peter Craven have been played after the frequency dependent weightings. The next item pair consists of pairwise panning and frequency dependent weighting. Afterwards, the weightings of Peter Craven [Cra03] have been opposed with those of Mark Poletti [Pol07]. In a last test, the latter weightings have been compared with the frequency dependent method. The arrangement of the item pairs is depicted in Table 5.1.

As expected, the listener did not know which kind of technique was used for sound creation. Every participant received a questionnaire with one sheet per angular range and four lines available per item pair to which he was allowed to listen up to two times. The questionnaire is included in the Sect. A.3.

Item	Sample 1	Sample 2
A	PenF	Craven
B	PW	PenF
C	Craven	Pen
D	Pen	PenF

Table 5.1: Assignment of the different weighting methods to the listening items.

5.3 Results

Angular range from $-45°$ to $45°$

Evaluation of the Records of the Questionnaire

The records of the listening test are attached in Sect. A.4. This section discusses the results and tries to find tendencies or characteristics. For the first sample pair in the angular range from $-45°$ to $45°$ there is a clear tendency towards the frequency-dependent panning function (PenF) regarding start and end position compared to the Craven sample. The panning seems to be more smoothly for PenF. One listener mentioned that the PenF sample could be more high frequent. For the second item B in the same angular range there is no clear tendency and both samples, PW and PenF, seem to perform similarly. Again, one participant notices coloration of the PenF sample and another one differences in amplitude. For the pair Craven versus PenF, the latter sample seems to start and end further out that is nearer to the start/end point $\pm 45°$. The PenF is respectively stuck to the outer speakers. Again, a difference in timbre is noticed even if both weighting methods are not frequency dependent. The last sample pair for the angular range $-45°$ to $45°$, Pen and PenF, appear to pan similarly and well from start to end point. PenF seems to be a little bit further out at large angles. Some difference in timbre is noticed and for both, high frequency roll-off between the physical speakers.

Discussion of the Results

Summarizing, one can say that the frequency-dependent method PenF performs at least equally well or slightly better for all of the item pairs. A difference in timbre was noticed several times even for the methods with constant frequency response. For that reason, one cannot deduce clear statements. One explanation for the difference in timbre can be the room reflections. Another explanation could be the bias of the listener from previously played samples, which possess partly similar panning methods and additionally frequency dependency.

Angular range from 110° to 250°

Evaluation of the Records of the Questionnaire

The area behind the listener starts again with the samples PenF versus Craven. There is no clear tendency towards one of the samples, both stick to the outer loudspeakers for a while and pan well directly behind the listener. One listener perceives a partly frontal image for the Craven sample. The image near the loudspeaker seems to be wider than directly behind the seat. For the second pair PW versus PenF, the records result in no clear tendency towards one method. One person mentions coloration for the PenF, another one frontal confusion for the pairwise panning method and other persons notice frontal confusion without specification of the sample. In general, the panning seems to be more smoothly than for the first item A and the images perform similar with slight angle differences. Similar to the previous items, the third sample pair sticks again to the loudspeakers for some time. The same person which noticed coloration for the pairs before, mentions it in this case for the non frequency-dependent methods. The notes reach from 'Both poor' to 'Smooth movement from start to finish for both samples', i.e., no clear tendency is obvious. Again, the last sample pair is stuck to loudspeakers for some time and then panned smoothly in between them. They seem to be hard to distinguish. Pen confuses frontal, while PenF is leading/better without a more profound description of the participants.

Discussion of the Results

In general one can say that the distance of the loudspeakers does not allow a perfect sound field synthesis. For that reason the images of the synthesized source are stuck to the speakers for some time. In the area in between, the panning results are satisfying for all samples without a clear favorite. Another feature which can be deduced is that the image seems to be wider near the loudspeaker. Besides the artifacts in sound field, the bad localization of the human hearing at angles left and right of a person [Bla97] could be a reason for that. Furthermore, especially one participant noticed frontal confusion several times, but never for the frequency-dependent method. Several others noticed also frontal confusion, but never named a specific sample. Furthermore, the sample pairs produce a bias in the perception if similar methods have been used, i.e., PenF consists of pairwise panning for high frequencies and the cosine penalty weighting for low frequencies, and so biases the perception if one of this methods is its sample partner.

Angular range from $30°$ to $110°$

Evaluation of the Records of the Questionnaire

The order of the items is the same for the source angles left of the listener. If one considers the records of the first two samples, it is obvious that PenF creates a much smoother panning between the two loudspeakers especially for the first degree steps. A perceived coloration is not mentioned, but both stick to the speakers again. The next samples PW and PenF show both a similar panning behavior between the speakers, whereby they are stuck again at the loudspeakers. Two of the five participants preferred the PenF method regarding smooth movement. For this case, colorations and timbre differences have been mentioned by several persons. At the next sample comparison, Craven's weighting seems to start between the front left and the center speaker and is stuck for a while at $30°$. In the end, both samples panned together. One listener again noticed coloration for this non coloration weighting methods. In the last comparison,

PenF has some slight benefits regarding smooth panning between the loudspeakers. Four persons named coloration or timbre differences for the PenF sample.

Discussion of the Results

Considering the evaluation, there is a slight tendency towards the frequency-dependent method, however some coloration has been mentioned, but this was also the case for the non frequency-dependent sample pair. The coloration can be also caused by sound reflections in the room. Again, this range is worse than frontal area where three speakers with low angular distance are available for the reproduction.

Angular range from 250° to 330°

Evaluation of the Records of the Questionnaire

The last angular range is symmetric to the one discussed before. It is difficult to recognize a tendency out of the records for the first sample comparison of PenF and Craven. Three persons assume that PenF departs quicker from the left surround speaker and Craven lags behind. Furthermore, a timbre difference for PenF is noticed of two participants, and another one perceives a reduced volume for PenF. If one considers the next comparison, the movement seems to be the same for both samples. Again, PenF depicts distortion features what has been mentioned by three listeners. Cravens sample states a smoother movement compared to Pen departing from 250°. Pen creates a timbre change near the end of the angular range which was mentioned by two persons. For comparison of Pen and PenF, it is conspicuous that there is no timbre difference mentioned. The movement of Pen seems to be quicker and more smoothly in this case.

Discussion of the Results

The last angular range creates a clear bias towards the first sample in each item. Several listeners described the first sample as the leading one. Another interesting thing is that for the first three items even for the one without frequency dependency,

a distortion was mentioned by several participants. One possible explanation for this is that especially on the right side of the listener, the reflections of the room have been particularly pronounced.

Conclusions from the Listening Test

The performed listening test is surely not a perfect proof for the performance of the included techniques. But it gives definitely an idea if more sophisticated investigations should be conducted. For almost all of the angular ranges, the PenF achieves positive results regarding smooth movement. Only in the last range, a bias towards the first sample in each item is shown. Furthermore, several times coloration has been mentioned. As the coloration was also mentioned for the frequency-independent sample pair, it is not possible to deduce a tendency towards one of the samples. The reflections of the room have certainly a big influence on this perception, too. Another outcome is the bias of the samples to each other: Besides the influence within an item, samples of previous items can influence the perception on the one currently played.

Chapter 6

Conclusions and Future Work

Within the scope of this thesis, solutions for the loudspeaker weightings for arbitrary reproduction setups have been derived based on the Ambisonics theory. The solutions have been given by matching the desired sound field with the sound field produced by the loudspeaker. A connection between pressure matching [Pol07] and mode matching [Pol00] has been shown. By the use of this matching condition, a direct solution for uniformly spaced positioned loudspeakers for the three and two dimensional case has been specified. The solutions for nonuniform layouts like the ITU layout are then deduced with a least squares approach as used in [Pol07] [Pol05] [BK10]. As this approach results in a non robust solution, the loudspeaker amplitudes are regulated with an angular penalty. Several penalizing techniques have been introduced and analyzed on the basis of the ITU layout loudspeaker driving functions. For the pairwise penalty, a connection to traditional stereophony has been derived. By the use of interpolation points which are defined by the limits of the Ambisonics theory regarding frequency and radius, a frequency-dependent weighting method has been introduced. To analyze the performance of the introduced techniques, an Ambisonics decoder with graphical user interface has been implemented. The tool provides controlling and analyzing elements for the design of the decoder, the reproduction setup control, audio demo control and test signal creation. With the help of this tool, a listening test has been performed and evaluated. Despite all that, one has to say that the results of the listening test are

too inconclusive to derive a firm statement. For that reason, further tests would be required.

The mentioned reverberant sound field leads to the conclusion that it would be necessary to perform a test in an Anechoic environment. Furthermore, a more sophisticated construction of the reproduction setup with additional loudspeaker for source positioning would lead to statistical significant results, like it has been done in [PH05]. Even though the test group of the listening test was rather small, the results give ideas if further investigations shall be done.

Based on the results presented in this thesis, further research could deal with the optimal solution for the frequency dependent weighting. One possibility would be to compare the head related transfer functions (HRTF) of the desired wave at a dummy head with one produced by the frequency-dependent method. The weightings could be optimized based on that. Furthermore, HRTF represent a useful evaluation method for the performance of a technique.

Appendix A

Appendix

A.1 Spherical Harmonics

This section shall give a brief overview how the spherical harmonics are defined.The m-th spherical harmonic of order n is defined as

$$Y_n^m(\theta, \phi) = \Lambda_n^m P_n^{|m|}(\cos(\theta)) e^{-\mathrm{i}m\phi}. \tag{A.1}$$

The dependency of elevation angle θ is given by the $|m|$th Legendre function of order n, which is defined as

$$P_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m}(P_n(x)) = \frac{(-1)^m}{2^n n!}(1-x^2)^{m/2} \frac{d^{n+m}}{dx^{n+m}}(x^2-1)^n, \tag{A.2}$$

where $x = \cos(\theta)$ is used in the spherical harmonics definition. The weighting factor Λ_n^m is defined as

$$\Lambda_n^m = \sqrt{\left(\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}\right)}. \tag{A.3}$$

In Table A.1, all spherical harmonics up to order $N = 3$ are shown. The angular dependencies are respectively given by the sine and cosine terms coming from the Legendre functions for the elevation angle θ and by the exponential term for the azimuthal angle

ϕ. It is obvious that following correspondence holds

$$Y_n^m(\theta,\phi)^* = Y_n^{-m}(\theta,\phi), \tag{A.4}$$

what creates a symmetry in the table.

$m\backslash n$	0	1	2	3
-3				$\sqrt{\frac{35}{64\pi}}\sin^3\theta e^{-3\mathrm{i}\phi}$
-2			$\sqrt{\frac{15}{32\pi}}\sin^2\theta e^{-2\mathrm{i}\phi}$	$\sqrt{\frac{105}{32\pi}}\sin^2\theta\cos\theta e^{-2\mathrm{i}\phi}$
-1		$\sqrt{\frac{3}{8\pi}}\sin\theta e^{-\mathrm{i}\phi}$	$\sqrt{\frac{15}{8\pi}}\sin\theta\ \cos\theta e^{-\mathrm{i}\phi}$	$\sqrt{\frac{21}{64\pi}}\sin\theta\,(5\cos^2\theta-1)\,e^{-\mathrm{i}\phi}$
0	$\sqrt{\frac{1}{4\pi}}$	$\sqrt{\frac{3}{4\pi}}\cos\theta$	$\sqrt{\frac{5}{16\pi}}\,(3\cos^2\theta-1)$	$\sqrt{\frac{7}{16\pi}}\,(5\cos^3\theta-3\cos\theta)$
1		$-\sqrt{\frac{3}{8\pi}}\sin\theta\,e^{\mathrm{i}\phi}$	$-\sqrt{\frac{15}{8\pi}}\sin\theta\ \cos\theta\,e^{\mathrm{i}\phi}$	$-\sqrt{\frac{21}{64\pi}}\sin\theta\,(5\cos^2\theta-1)\ e^{\mathrm{i}\phi}$
2			$\sqrt{\frac{15}{32\pi}}\sin^2\theta\,e^{2\mathrm{i}\phi}$	$\sqrt{\frac{105}{32\pi}}\sin^2\theta\cos\theta\,e^{2\mathrm{i}\phi}$
3				$-\sqrt{\frac{35}{64\pi}}\sin^3\theta\,e^{3\mathrm{i}\phi}$

Table A.1: Spherical Harmonics up to order $N = 3$.

Fig. A.1 depicts the normalized real part of the spherical harmonics up to order $N = 3$. As the shape of the harmonics shows symmetrical correspondence, it has been plotted only for positive values of m.

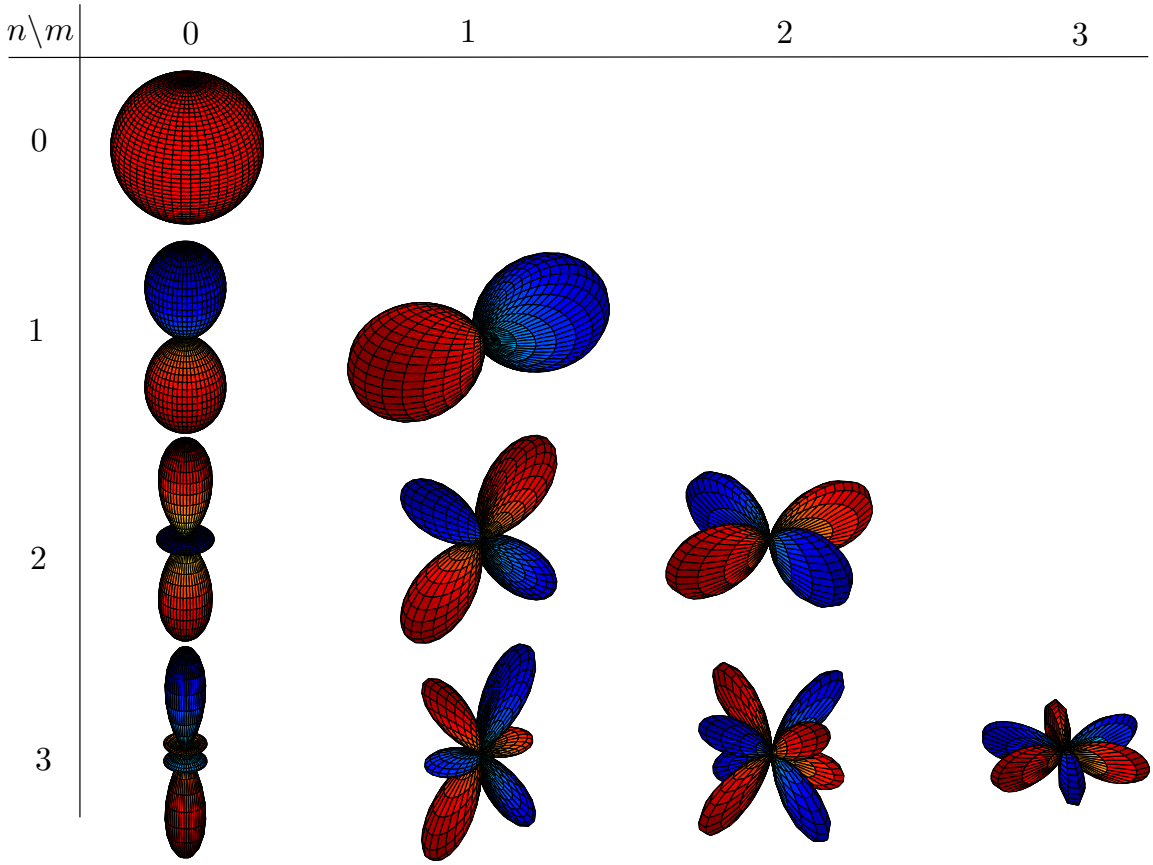

Figure A.1: Real part of the spherical harmonics up to order $N = 3$.

A.2 2D- and 3D- Greens Function

The Green's function represents a solution for the wave equation and describes the behavior of a radiating source. For a line source the solution is given by [Wil99]

$$G_{2D}(\|\mathbf{r} - \mathbf{r}_0\|) = \frac{\mathrm{i}}{4} H_0^{(1)}(k \, \|\mathbf{r} - \mathbf{r}_0\|), \tag{A.5}$$

whereby $H_0^{(1)}$ represents the Bessel function of the third kind of first order also called Hankel function. The Hankel function consists of the Bessel function of the first kind $J_\nu(x)$ and second kind $Y_\nu(x)$ also known as Weber or Neumann functions [Wil99]. The connection is given by

$$H_\nu^{(1)}(x) = J_\nu(x) + \mathbb{i} \cdot Y_\nu(x), H_\nu^{(2)}(x) = J_\nu(x) - \mathbb{i} \cdot Y_\nu(x). \tag{A.6}$$

For an infinite length there is no dependency of z for a line source. A discussion about line and point sources in a more practical way is done in section 2.3. For a point source, the solution of the wave equation is given by the three dimensional Green's function

$$G_{3D}(\|\mathbf{r} - \mathbf{r}_0\|) = \frac{e^{\mathbb{i}k\|\mathbf{r}-\mathbf{r}_l\|)}}{4\pi \|\mathbf{r} - \mathbf{r}_0\|} = \frac{\mathbb{i}}{4\pi} h_0^{(1)}(k \|\mathbf{r} - \mathbf{r}_0\|). \tag{A.7}$$

For the spherical Hankel function $h_0^{(1)}$, the coherence of A.6 holds, too.

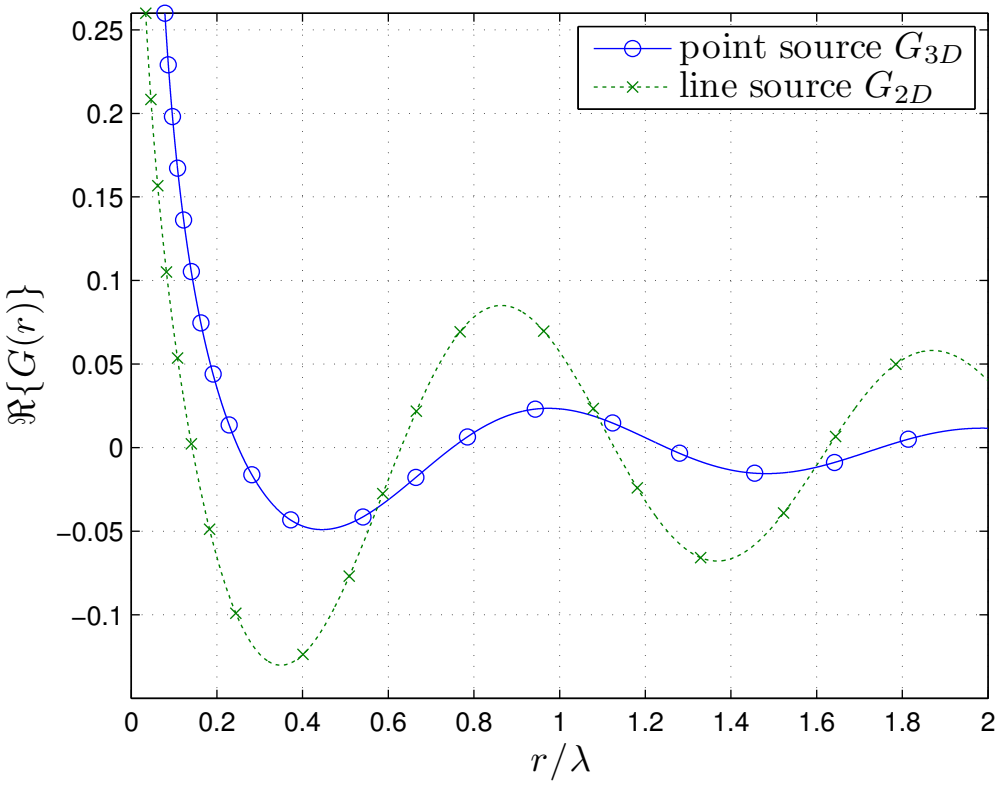

Figure A.2: Comparison of the real part of the Green's function for a spherical and a cylindrical wave.

In figure A.2 the real part of the two and three dimensional Green's function is depicted. It seems that for values of radius r near to zero, the point source possesses a higher amplitude. For the spherical wave, the imaginary part simplifies to

$$\Re\left\{G_{3D}\right\} = \Re\left\{\frac{e^{\mathrm{i}k\|\mathbf{r}-\mathbf{r}_l\|)}}{4\pi\left\|\mathbf{r}-\mathbf{r}_0\right\|}\right\} = \Re\left\{\frac{\mathrm{i}}{4\pi}h_0^{(1)}(k\left\|\mathbf{r}-\mathbf{r}_0\right\|)\right\} = -y_0(r) = \frac{\cos(r)}{r}. \quad \text{(A.8)}$$

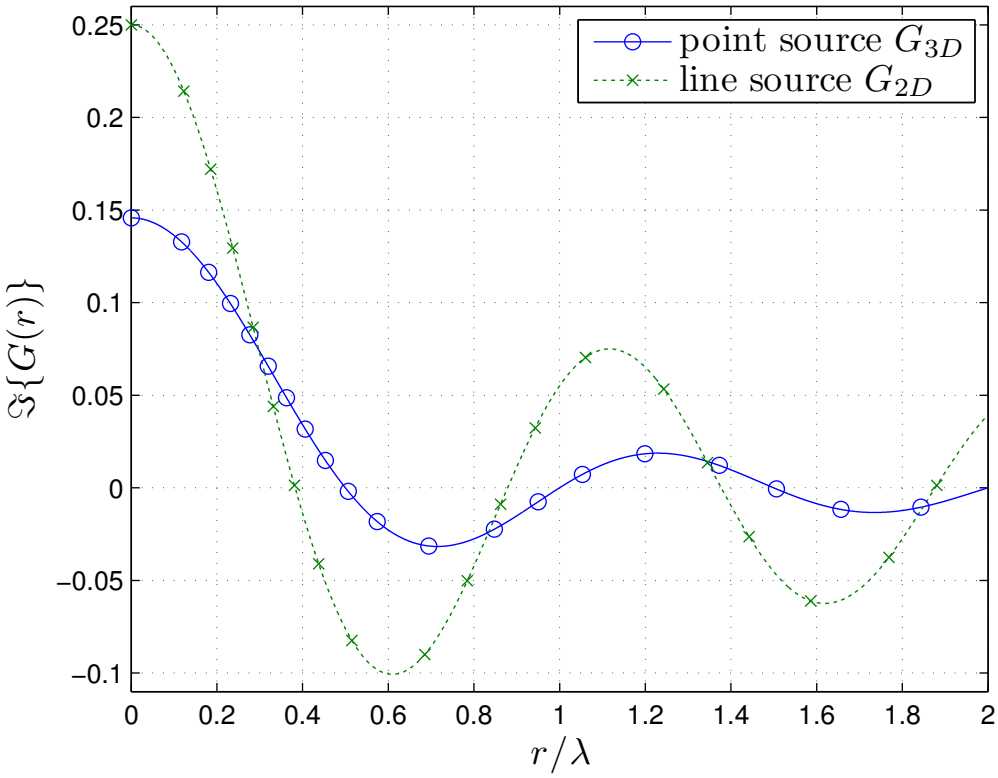

Figure A.3: Comparison of the imaginary part of the Green's function for a spherical and a cylindrical wave.

Figure A.3 shows the plot of the imaginary of a spherical wave. In this case, the line source begins with an higher value and attenuates more slowly. The imaginary part of quation A.7 results in the spherical Bessel function

$$\Im\left\{G_{3D}\right\} = \Im\left\{\frac{e^{\mathrm{i}k\|\mathbf{r}-\mathbf{r}_l\|)}}{4\pi\left\|\mathbf{r}-\mathbf{r}_0\right\|}\right\} = \Im\left\{\frac{\mathrm{i}}{4\pi}h_0^{(1)}(k\left\|\mathbf{r}-\mathbf{r}_0\right\|)\right\} = j_0(r) = \frac{\sin(r)}{r}, \quad \text{(A.9)}$$

which is the cardinal sine.

As one can see from the real part of the Green's function, the point source attains counter-intuitively higher values for small radii. Even if this does not hold for the

imaginary part, it would be interesting to know how the magnitude of the two Green's functions behaves. Fig. A.4 shows the plot for both versions. The magnitude of the two dimensional Green's function is proportional $1/r$ as one can see from Eq. A.7. It obvious that both curves intersect for $r \approx 0.1$. The previous plot does not show exactly

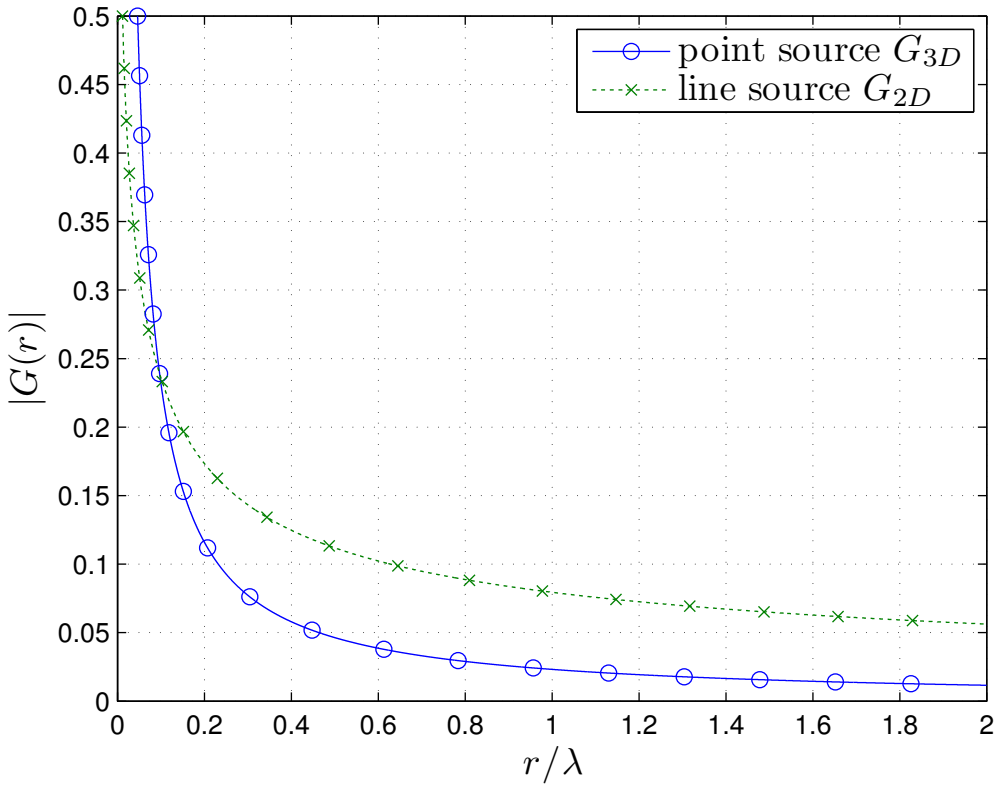

Figure A.4: Comparison of the magnitude of the Green's function for a spherical and a cylindrical wave.

which attenuation the two dimensional Green's function possesses. A logarithmic plot is more meaningful in that case. Fig. A.5 depicts the magnitude in Decibel of the functions in logarithmic scaling. Considering the slope of the curve the attenuation for the spherical wave is 20dB per decade what is not surprising because the proportionality to $1/r$. The slope for the magnitude of the two dimensional Green's function attains -10dB per decade what leads to a magnitude proportional to $1/\sqrt{r}$. Summarizing one can say, that an infinite long line source has the half attenuation of a point source.

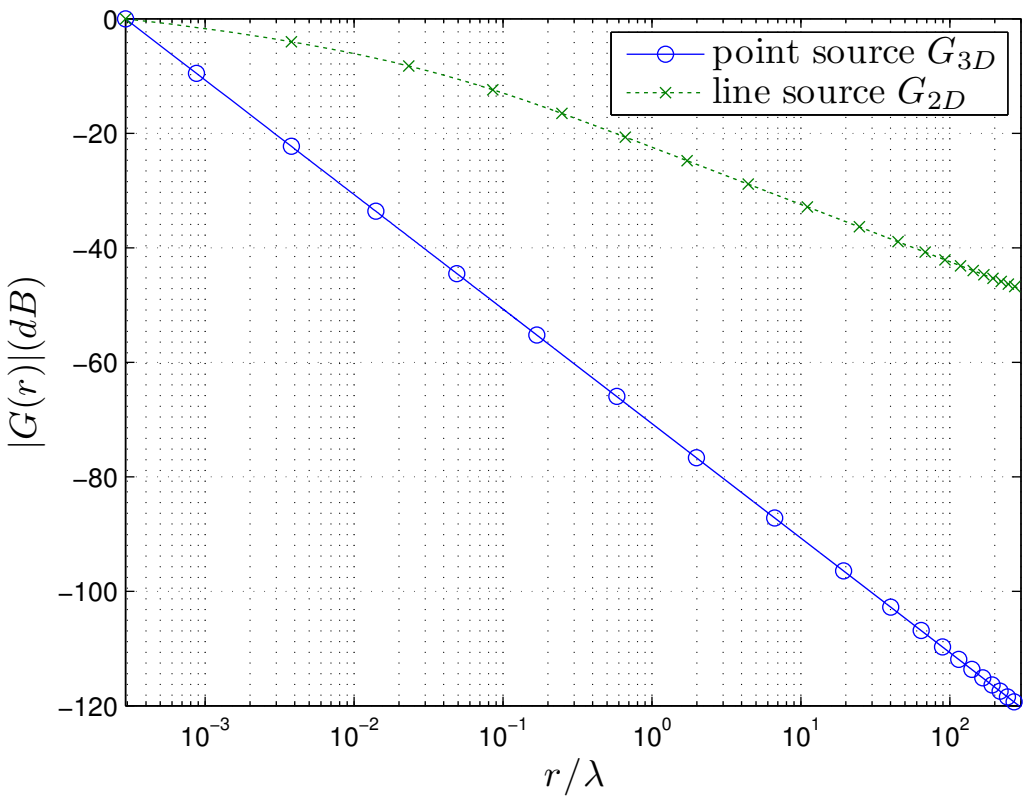

Figure A.5: Comparison of the magnitude in logarithmic scaling of the Green's function for a spherical and a cylindrical wave.

A.3 Questionnaire Listening Test

Source Movement for $\phi_s = -45°...45°$ in $5°$ steps

Item A1 (First Sound) vs A2 (Second Sound): Notes

Item B1 (First Sound) vs B2 (Second Sound): Notes

Item C1 (First Sound) vs C2 (Second Sound): Notes

Item D1 (First Sound) vs D2 (Second Sound): Notes

Source Movement for $\phi_s = 110°...250°$ in $5°$ steps

Item A1 (First Sound) vs A2 (Second Sound): Notes

Item B1 (First Sound) vs B2 (Second Sound): Notes

Item C1 (First Sound) vs C2 (Second Sound): Notes

Item D1 (First Sound) vs D2 (Second Sound): Notes

Source Movement for $\phi_s = 30°...110°$ in $5°$ steps

Item A1 (First Sound) vs A2 (Second Sound): Notes

Item B1 (First Sound) vs B2 (Second Sound): Notes

Item C1 (First Sound) vs C2 (Second Sound): Notes

Item D1 (First Sound) vs D2 (Second Sound): Notes

Source Movement for $\phi_s = 250°...330°$ in $5°$ steps

Item A1 (First Sound) vs A2 (Second Sound): Notes

Item B1 (First Sound) vs B2 (Second Sound): Notes

Item C1 (First Sound) vs C2 (Second Sound): Notes

Item D1 (First Sound) vs D2 (Second Sound): Notes

A.4 Notes Records of Questionnaire

Item A1 vs Item A2 with $\phi_s = -45°...45°$

Listener 1

A1 moved slower from $-45°$ to $0°$ and A2 seemed to pan more evenly. A1 made a fast jump to the centre speaker. But then A1 seemed to create a more convincing (clearer) image from $0°$ to $45°$ whilst A2 lagged behind.

Listener 2

A2 seems to be stuck on the speakers.
A pans quite well

Listener 3

A1 - Initially A2 was following A1 later it got reversed. There was a jump at the beginning.

Listener 4

A1 much further out than A2; A1 better; slight diff in timbre - maybe A1 slightly more high freq?

Listener 5

A2 sounded more central when position was in between physical speakers, but positions were equivalent at speakers. No difference in colouration;

Item B1 vs Item B2 with $\phi_s = -45°...45°$

Listener 1

B1 & B2 seemed to perform similarly. They both 'stuck' to the front-right loudspeaker for a little long. Perhaps B1 sounded a little more convincing, but I would have to listen to the pair of sounds in the opposite order to be sure. Images were both clear;

Listener 2

Both pan well, B2 seems to be slightly distorted at large angle ($\approx 40+°$)

Listener 3

Both went together. But amplitude was higher few time on 1st.

Listener 4

B1 slightly further out at 45° than B2.

Listener 5

Both sources appeared to come from the same locations. No difference in colouration.

Item C1 vs Item C2 with $\phi_s = -45°...45°$

Listener 1

C1 moved faster than C2 which appeared to be stuck to the FR loudspeaker initially. So C1 led C2 upto the front-left (FL) loudspeaker, but then C2 overtook C1, eventually localizing at 45° whilst C1 ended at around 30°. Images were fairly clear.

Listener 2

C2 seems to be slightly further behind (larger angle)

Listener 3

Both seem to come from different angle. This is more so in the beginning and at the end. The first sound was advanced.

Listener 4

C2 further out eg at 45° than C1. Difference in timbre.

Listener 5

C2 sounded distorted - missing low frequencies. C2 appeared to cling to physical speakers while C1 moved smoothly between.

Item D1 vs Item D2 with $\phi_s = -45°...45°$

Listener 1

Again D2 lagged D1. D1 was perhaps more clear. I would like to hear the sound in opposite order (D2 first) to confirm.

Listener 2

Both pan well, D2 slightly better at large angles

Listener 3

Both have constant movement and together.

Listener 4

No significant difference in source angle. Slight timbre difference.

Listener 5

Both sources appeared to move smothly between loudspeakers. Also, both sources seemd to have high-frequency roll-off between physical speakers.

Item A1 vs Item A2 with $\phi_s = 110°...250°$

Listener 1

A2 led A1 perhaps slightly, fairly clear image, sticking to back-left speaker a little at first but not too much. Quite similar in performance. Stuck also a bit at back-right speaker.

Listener 2

Both pan well behind the seat, sound is stuck near the 110°/250° points.

Listener 3

Both were together. But both stayed longer time in the side speakers.

Listener 4

A1 is to the rear. A2 is confusing and partly frontal. A1 is good

Listener 5

No difference in colouration between A1/A2. Both sources appeared to spreadout before moving behind me and then again before settling at destination.

Item B1 vs Item B2 with $\phi_s = 110°...250°$

Listener 1

Quite wide images, B2 leading, especially behind me. Colouration of signals? Esp. the second one? Images perhaps harder to distinguish than in prior cases.

Listener 2

Both do well at different angles. B2 sounds like it is at a slightly different angle to B1.

Listener 3

Bothe travelling same way. Stayed little long on the right speaker. Both appeared to come from front at the half point.

Listener 4

B2 is to the rear. B1 is confusing / frontal

Listener 5

No difference in colouration. Some apparent front/back confusion when slightly of physical speaker, but much less apparent than test A above.

Item C1 vs Item C2 with $\phi_s = 110°...250°$

Listener 1

Wider images again, jumping quite quickly from back-left speaker to behind, but seemingly fairly smoothly from behind to back-rigth. C2 leading C1 as the source was panned. Slight colouration as in item B.

Listener 2

C2 sticks to the rear speakers.

Listener 3

Both stayed together, but later part it appeared from the front speaker.

Listener 4

Both poor: C1 slightly move to the rear.

Listener 5

No difference in colouration. Smooth movement from start to finish for both samples.

Item D1 vs Item D2 with $\phi_s = 110°...250°$

Listener 1

Both stuck on the loudspeaker back left. Then image quickly widening to move behind. Hard to distinguish the two different signals, but perhaps the second sound was leading again.

Listener 2

Both sound identical, both are good at resolving angle

Listener 3

Both stayed together, but long time on the right speaker

Listener 4

D2 better. D1 is confusing frontal.

Listener 5

No difference in colouration. No difference in position. Samples seemed to cling to speakers, but move smoothly between them

Item A1 vs Item A2 with $\phi_s = 30°...110°$

Listener 1

A1 leading A2 with better image quality and less likely to be stuck to loudspeaker.

Listener 2

A2 sticks to the front left speaker and is slightly different sounding to A1

Listener 3

1st in advanced initially but later together.

Listener 4

A1 starts further round 30°. Both end up similar at 110°

Listener 5

No difference in colouration. A1 appeared to move between physical speakers more smoothly than A2. A2 seemed to cling to speakers.

Item B1 vs Item B2 with $\phi_s = 30°...110°$

Listener 1

B2 leading B1 at first. B2 panning more smoothly whilst B1 jumped. B2 seemed to be 'coloured' differently, which made me feel, this image was wider.

Listener 2

Both pan well, some attenuation while the source moves between the front and rear speakers.

Listener 3

B1 stayed on the same speaker for longer time, later caught up.

Listener 4

Similar. Both slow to move. Stuck at 30° for a few cases. Slight timbred differences. Source angles similar.

Listener 5

Near 30°, B1 appeared to have some high-frequency roll-off. Both samples appeared to move smoothly between speakers.

Item C1 vs Item C2 with $\phi_s = 30°...110°$

Listener 1

C1 leading C2. C1 seemed to start between the front-left and the centre loudspeakers but quickly moved to the FL loudspeaker. C2 started at FL loudspeaker.

Listener 2

C1 sticks to the front speaker at the beginning of the test, otherwise both are good at panning.

Listener 3

Both stayed long on the speakers. Both moved equally.

Listener 4

C2 further round than C1. C1 stuck at 30° . End up save at 110°

Listener 5

C2 appeared to have a slight bias toward 90°. There was a subtle difference in colouration, but difficult to describe.

Item D1 vs Item D2 with $\phi_s = 30°...110°$

Listener 1

Similar performance. Both were fairly 'sticky' front-left loudspeaker, but still panned quite gradually through from 30° to 110°. D2 started to move from front-left first, with D1 following initially. Perhaps D1 moved faster though angle once unsticking from front-left loudspeaker

Listener 2

D2 pans better, but is slightly distorted (low-pass filter?)

Listener 3

Both jumped from speaker to another

Listener 4

Very similar source angles. D1 higher freq sp. for angles $\approx 45°$

Listener 5

D2 had high-frequency roll of when in between physical speakers. Both samples moved smoothly over range.

Item A1 vs Item A2 with $\phi_s = 250°...330°$

Listener 1

A2 followed A1 through to 330°, but A2 moved through front right loudspeaker to the centre one. It took a little while fpr the images to move from back right.

Listener 2

A1 is quieter than A2 as the source approaches 330°

Listener 3

A2 was behind A1, moving in jumpily.

Listener 4

A1 moves away from right surround quicker. A1 loses low frequencies as we approach 330°. Clear timbred diff- A1 timbre varied, A2 timbre constant.

Listener 5

A1 had high-frequency roll-off. Both samples appeared quieter around 330°. Both samples moved together, but appeared static around 250° and then jumped to around 300°.

Item B1 vs Item B2 with $\phi_s = 250°...330°$

Listener 1

B1 & B2 spent more time stuck than in item A. Quite a fast jump from back-right to front-right, with B1 leading B2. The image jumped the most rapidly through lateral angle 270°.

Listener 2

B2 sticks to the rear speaker angle, sounds distorted compared with B1.

Listener 3

Both stayed together, but longer in the later speaker.

Listener 4

Both stuck at 250° for a while. Rapid shift through 270°. B2 timbre changed with angle.

Listener 5

Both samples moved together, but again, movement was hard to detect at first. Colouration was normal until around 300°, then B2 suffered some low-frequency roll-off.

Item C1 vs Item C2 with $\phi_s = 250°...330°$

Listener 1

C1 panned more smoothly from back-right to front-right. C2 jumped quite fast to front-right.

Listener 2

Both pan well, as the source approaches 330°, but don't do well otherwise.

Listener 3

Both looking good, constant movement and together.

Listener 4

Same speaker detent at 250° as B1 B2. C2 timbre changed as -¿ 330° -¿ high-pass. C1 slightly better at getting away from 250°

Listener 5

C2 sounded band-passed near the end of the range. Both samples moved together, smoothly through the range.

Item D1 vs Item D2 with $\phi_s = 250°...330°$

Listener 1

D1 & D2 stuck to back right, with D1 leading D2. D1 creating a clearer image and panned more smoothly.

Listener 2

Both pan ok, D2 seems to be slightly offset (follows D1).

Listener 3

Constant movement and together.

Listener 4

D1 moves forward quicker (slightly) than D2, similar timbre.

Listener 5

Not colouration noticed. Movement started smooth but D1 progressed faster than D2 near 300° to completion.

Appendix B

Abbreviations

ANSI	American National Standards Institute
DTFT	discrete-time Fourier transform
GUI	graphical user interface
LS	left surround
HOA	higher order Ambisonics
HRTF	head related transfer functions
IFFT	inverse fast fourier transformation
IR	impulse response
ITU	International Telecommunication Union
PA	public address
pchip	piecewise cubic Hermite interpolating polynomial
rms	root mean squared
RS	Right Surround
VBAP	vector based amplitude panning
WFS	wave field synthesis
XML	extensible markup language

Appendix C

Notation

Mathematical Operators

$\frac{\partial}{\partial \mathbf{w}}$	derivative with respect to $\mathbf{w}$
$(\cdot)^*$	complex conjugation
$(\cdot)^{\mathbb{H}}$	complex conjugate transposition
$\lVert\cdot\rVert$	L^2 norm
$\Im\{\cdot\}$	imaginary part operator
$\Re\{\cdot\}$	real part operator
$(\cdot)^{\mathbb{T}}$	transposition

Symbols

a	piston radius
a_0	filter coefficient
a_M^L	Mth filter coefficient of the Lth loudspeaker
α	angular distance between loudspeakers
A_a	acoustic absorption
A_n^m	Ambisonics coefficient

$\mathbf{A}_{cy}^{ls}$	weighting matrix for matching a point source in cylindrical coordinates
$\mathbf{A}_{cy}^{pw}$	weighting matrix for matching a plane wave in cylindrical coordinates
$\mathbf{A}_{sp}^{ps}$	weighting matrix for matching a point source in spherical coordinates
$\mathbf{A}_{sp}^{pw}$	weighting matrix for matching a plane wave in spherical coordinates
$\mathbf{B}$	phase term matrix
c	speed of sound
D	directivity factor
d_c	chord distance
f	frequency
f_s	sampling frequency
$g[k]$	filter impulse response
$G(e^{\mathrm{i}\Omega})$	frequency response
$G_{2D}()$	two dimensional Green's function
$G_{3D}()$	three dimensional Green's function
γ_r	regularistion factor
$\mathbf{\Gamma}$	regularistion matrix
$h_n()$	spherical Hankel function of the first kind of n-th order
$H_m()$	Hankel function of the first kind of m-th order
$\mathbf{H}$	propagation matrix including the matching conditions
$\mathbf{H}_{cy}^{ls}$	propagation matrix including the matching conditions for a point source in cylindrical co
$\mathbf{H}_{cy}^{pw}$	propagation matrix including the matching conditions for a plane wave in cylindrical coor
$\mathbf{H}_{sp}^{ps}$	propagation matrix including the matching conditions for a point source in spherical coor
$\mathbf{H}_{sp}^{pw}$	propagation matrix including the matching conditions for a plane wave in spherical coord
i	imaginary unit $\sqrt{-1}$
I_d	direct sound intensity
I_r	reverberant sound intensity
$j_n()$	spherical Bessel function of the first kind of n-th order
$J_n()$	Bessel function of the first kind of n-th order
k	wave number

$\mathbf{k}$	wave vector
K	number of frequency points
ω	angular frequency
Ω	normalized angular frequency
Ω_μ	μ-th normalized angular frequency
p	sound pressure
$\mathbf{p}$	mode or pressure matching vector
P_a	acoustic power
$P_n^{\lvert m\rvert}$	$\lvert m\rvert$-th associated Legendre function of order n
ϕ	azimuthal angle
ϕ_l	azimuthal angular position of the loudspeaker
ϕ_s	azimuthal angular position of the sound source
ϕ_p	azimuthal angular position of point p
$\mathbf{Q}$	trigonometric coefficient matrix
r	distance to the origin in spherical coordinates (or radius)
R	distance to the origin in cylindrical coordinates
$\mathbf{r}$	position vector
$\mathbf{r}_0$	position vector of a source
$\mathbf{r}_p$	position vector of point p
T_{60}	reverberation time
$\mathbf{T}$	trigonometric function matrix
θ	elevation angle
θ_l	elevation angular position of the loudspeaker
θ_s	elevation angular position of the sound source
θ_p	elevation angular position of point p
V	volume
w_l	weighting of the l-th loudspeaker
$\mathbf{w}$	weighting vector with elements w_l
$\mathbf{W}$	weighting vector with elements w_l

Y_n^m	m-th spherical harmonic of n-th order

List of Figures

2.1 Spherical coordinates definition scheme. 6

2.2 Modes to match given by the sectorial approximation marked with ■. Modes which are zero for $\theta = \pi/2$ marked with ∘. Left out modes marked with □ (adapted from [PBA12]). 11

2.3 Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 29 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 1000$Hz. 12

2.4 Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 55 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 1000$Hz. 13

2.5 Snapshot of sound field synthesized with HOA using a circular distribution ($r = 1.5$m) of 55 spherical point sources. The virtual source constitutes a plane wave with an incidence angle of 0° and frequency $f_{pw} = 2000$Hz. 13

2.6 Panning function on a sphere depending on azimuthal angle ϕ_s, elevation angle θ_s of the source, for $L = 25$ loudspeaker. 20

2.7 Cylindrical coordinates definition scheme. 21

2.8 Panning function depending on the source angle ϕ_s for loudspeaker placed at $\phi_l = 0$ with a reproduction setup of equally distributed loudspeakers placed on circle. 28

3.1 ITU reproduction layout as described in [IR12]. 36
3.2 Interpolants for the ITU loudspeaker layout. 37
3.3 Plot of several penalty functions with low angular distance in the center and high angular distance on the left/right side. 39
3.4 Panning functions for the ITU layout regularized by $\gamma_r = 1.5$ with the cosine penalty function. Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights. . . 41
3.5 Panning functions for the ITU layout regularized by $\gamma_r = 0.15$ with the exponential penalty function ($b = 4, p = 1$). Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights. 42
3.6 Panning functions for the ITU layout regularized by $\gamma_r = 0.5$ with the exponential penalty function ($b = 4, p = 4$). Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights. 42
3.7 Panning functions for the ITU layout regularized by $\gamma_r = 2$ with the pairwise penalty function. Unlimited solution on the left, limited to an order of 4 on the right. − − − rms power, − · − · − sum of weights. . . 43
3.8 Sine-, mode and tangent panning law in comparison for loudspeakers at $\phi_l = \pm 30°$. 46
3.9 Plot of two different interpolation methods for four defined points. . . . 48
3.10 Exemplary plot of an impulse response with 64 samples. 49

4.1 Window for the creation of an interpolation point. 54
4.2 Interface for the control of the decoder and the listening test. 55
4.3 Interface for applying a filter on the created signal. 56

5.1 Picture of the room with reproduction setup and a dummy head on the position of the listener. 58

A.1 Real part of the spherical harmonics up to order $N = 3$. 71

A.2 Comparison of the real part of the Green's function for a spherical and a cylindrical wave. 72

A.3 Comparison of the imaginary part of the Green's function for a spherical and a cylindrical wave. 73

A.4 Comparison of the magnitude of the Green's function for a spherical and a cylindrical wave. 74

A.5 Comparison of the magnitude in logarithmic scaling of the Green's function for a spherical and a cylindrical wave. 75

List of Tables

5.1 Assignment of the different weighting methods to the listening items. . 62

A.1 Spherical Harmonics up to order $N = 3$. 70

Bibliography

[AS00] Ahnert, W. ; Steffen, F.: *Sound Reinforcement Engineering: Fundamentals and Practice*. Taylor & Francis, 2000 http://books.google.com.au/books?id=GD5lGfU3UQgC. – ISBN 9780415238700

[AS08a] Ahrens, J. ; Spors, S.: Analytical driving functions for higher order Ambisonics. In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (2008). http://dx.doi.org/10.1109/ICASSP.2008.4517624. – DOI 10.1109/ICASSP.2008.4517624. – ISBN 978–1–4244–1483–3

[AS08b] Ahrens, Jens ; Spors, Sascha: An Analytical Approach to Sound Field Reproduction using Circular and Spherical Loudspeaker Distributions. In: *Acta Acoustica united with Acoustica* 94 (2008), November/December, Nr. 6, 988–999. https://laboratories-afi.telekom.de/PubDB/Documents/AhrensSpors_ActaAcustica94.pdf. – ISSN 1610–1928

[Bau62] Bauer, B.: Phasor analysis of some stereophonic phenomena. In: *IRE Transactions on Audio* 10 (1962). http://dx.doi.org/10.1109/TAU.1962.1161613. – DOI 10.1109/TAU.1962.1161613. – ISSN 0096–1981

[BD98] Brown, C.P. ; Duda, R.O.: A structural model for binaural sound synthesis. In: *IEEE Transactions on Speech and Audio Processing* 6

(1998), Nr. 5, 476–488. http://dx.doi.org/10.1109/89.709673. – DOI 10.1109/89.709673. – ISSN 10636676

[Ber54] Beranek, L.L.: *Acoustics*. McGraw-Hill, 1954 (McGraw-Hill electrical and electronic engineering series). http://books.google.de/books?id=Czm1AAAAIAAJ

[Ber73] Bernfeld, Benjamin: Attempts for Better Understanding of the Directional Stereophonic Listening Mechanism. In: *Audio Engineering Society Convention 44*, 1973

[BK10] Batke, Johann-markus ; Keiler, Florian: Investigation of Robust Panning Functions. (2010)

[Bla97] Blauert, J.: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997 http://books.google.com.au/books?id=wBiEKPhw7r0C. – ISBN 9780262024136

[BV10] Boyd, Stephen ; Vandenberghe, Lieven ; Press, Cambridge U. (Hrsg.): *Convex Optimization*. Cambridge University Press, 2010. – 487–487 S. http://dx.doi.org/10.1080/10556781003625177. http://dx.doi.org/10.1080/10556781003625177. – ISBN 9780521833783

[CK12] Colton, D. ; Kress, R.: *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer New York, 2012 (Applied Mathematical Sciences). http://books.google.co.nz/books?id=sLFc8Jwdm9UC. – ISBN 9781461449416

[Cra03] Craven, Peter G.: Continuous Surround Panning for 5-Speaker Reproduction. In: *24th Internation Audio Engineering Society Conference on Multichannel Audio* (2003), S. 1–6

[Dan03] Daniel, Jérôme: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In: *AES*. Copenhagen, 2003, 1–15

[Ger73] Gerzon, Michael A.: Periphony: With-Height Sound Reproduction. In: *J. Audio Eng. Soc* 21 (1973), Nr. 1, 2–10. http://www.aes.org/e-lib/browse.cfm?elib=2012

[Ger85] Gerzon, Michael A.: Ambisonics in Multichannel Broadcasting and Video *. In: *Journal of the Audio Engineering Society* 33 (1985), 859–871. http://sites.google.com/site/mytemporarydownloads/Gerzon_JAES_1985.pdf. – ISSN 00047554

[Haa51] Haas, Helmut: Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache [On the influence of a single echo on the audibility of speech]. In: *Acustica* 1 (1951), S. 49–58

[Hag89] Hager, William W.: Updating the Inverse of a Matrix. In: *SIAM Review* 31 (1989), Nr. 2, S. 221–239. http://dx.doi.org/10.1137/1031049. – DOI 10.1137/1031049. – ISBN 00361445

[Hjø11] Hjørungnes, A.: *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, 2011 http://books.google.co.nz/books?id=8abnhvLAyNAC. – ISBN 9781139498043

[IR12] ITU-R: Multichannel sound technology in home and broadcasting applications BS Series. In: *ITU-R Report* 4 (2012)

[J. 01] J. Daniel: *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Université Paris 6, Diss., 2001

[KFCS99] Kinsler, Lawrence E. ; Frey, Austin R. ; Coppens, Alan B. ; Sanders, James V.: *Fundamentals of acoustics*. Bd. 1. 1999. – 560 S. – ISBN 0–471–84789–5

[KMNF89] Kahaner, D. ; Moler, C.B. ; Nash, S. ; Forsythe, G.E.: *Numerical methods and software*. Prentice Hall, 1989 (Prentice-Hall series in computational mathematics). http://books.google.co.nz/books?id=jipEAQAAIAAJ

[KSAJ07] Kennedy, Rodney a. ; Sadeghi, Parastoo ; Abhayapala, Thushara D. ; Jones, Haley M.: Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields. In: *IEEE Transactions on Signal Processing* 55 (2007), Juni, Nr. 6, 2542–2556. http://dx.doi.org/10.1109/TSP.2007.893738. – DOI 10.1109/TSP.2007.893738. – ISSN 1053–587X

[Kuh77] Kuhn, George F.: Model for the interaural time differences in the azimuthal plane. In: *The Journal of the Acoustical Society of America* (1977)

[LCYG99] Litovsky, Ruth Y. ; Colburn, H. S. ; Yost, William A. ; Guzman, Sandra J.: The precedence effect. In: *The Journal of the Acoustical Society of America* 106 (1999), Nr. 4, 1633. http://dx.doi.org/10.1121/1.427914. – DOI 10.1121/1.427914. – ISSN 00014966

[LS03] Lee, Seung-Rae Lee Seung-Rae ; Sung, Koeng-Mo Sung Koeng-Mo: Generalized encoding and decoding functions for a cylindrical ambisonic sound system. In: *IEEE Signal Processing Letters* 10 (2003). http://dx.doi.org/10.1109/LSP.2002.806703. – DOI 10.1109/LSP.2002.806703. – ISSN 1070–9908

[ME04] Margolis, E. ; Eldar, Y.C.: Reconstruction of nonuniformly sampled periodic signals: algorithms and stability analysis. In: *Electronics, Circuits and Systems, 2004. ICECS 2004. Proceedings of the 2004 11th IEEE International Conference on*, 2004, S. 555–558

[Mol13] Moler, Cleve (.: *Numerical Computing with MATLAB*. Siam, 2013. – Chapter 3 S. `http://www.mathworks.com/moler/interp.pdf`

[Moo09] Moore, John D.: *The Development of a Design Tool for 5-Speaker Surround Sound Decoders*, Diss., 2009. – 287 S.

[OSB99] Oppenheim, Alan V. ; Schafer, Ronald W. ; Buck, John R.: *Discrete Time Signal Processing*. Bd. 1999. 1999. – 870 S. – ISBN 0137549202

[PBA12] Poletti, M.A. ; Betlehem, T. ; Abhayapala, T.D.: Analysis of 2D sound reproduction with fixed-directivity loudspeakers. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2012. – ISBN 9781467300469, 377-380

[PFN10] Poletti, M ; Fazi, F M. ; Nelson, P A.: Sound-field reproduction systems using fixed-directivity loudspeakers. In: *The Journal of the Acoustical Society of America* 127 (2010), S. 3590–3601. `http://dx.doi.org/10.1121/1.3409486`. – DOI 10.1121/1.3409486. – ISSN 00014966

[PH05] Pulkki, V. ; Hirvonen, T.: Localization of virtual sources in multichannel audio reproduction. In: *IEEE Transactions on Speech and Audio Processing* 13 (2005), Nr. 1. `http://dx.doi.org/10.1109/TSA.2004.838533`. – DOI 10.1109/TSA.2004.838533. – ISBN 1063–6676

[Pol96] Poletti, M.: The design of encoding functions for stereophonic and polyphonic sound systems. In: *Journal of the Audio Engineering Society* 44 (1996), Nr. 11, S. 948–963. – ISSN 1549–4950

[Pol00] Poletti, Mark A.: A Unified Theory of Horizontal Holographic Sound Systems. In: *Journal of the Audio Engineering Society* 48 (2000), 1155—-1182. http://www.aes.org/e-lib/browse.cfm?elib=12033

[Pol05] Poletti, MA: Three-Dimensional Surround Sound Systems Based on Spherical Harmonics*. In: *Journal of the Audio Engineering Society* 53 (2005), S. 1004–1025. ISBN 1549–4950

[Pol07] Poletti, Mark A.: Robust Two-Dimensional Surround Sound Reproduction for Nonuniform Loudspeaker Layouts. In: *Journal of the Audio Engineering Society* 55 (2007), S. 598–610

[PT55] Penrose, R. ; Todd, J. A.: A generalized inverse for matrices. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51 (1955), Nr. July, S. 406. http://dx.doi.org/10.1017/S0305004100030401. – DOI 10.1017/S0305004100030401. – ISBN 0305004100030

[Pul97] Pulkki, Ville: Virtual sound source positioning using vector base amplitude panning. In: *Journal of the Audio Engineering Society* 45 (1997), 456–466. http://www.aes.org/e-lib/browse.cfm?elib=7853. – ISSN 15494950

[RS09] Rabenstein, Rudolf ; Spors, Sascha: Sound Field Reproduction. Version: 2009. http://dx.doi.org/10.1121/1.3203918. In: Benesty, Jacob (Hrsg.) ; Sondhi, Mohan M. (Hrsg.) ; Huang, Yiteng (Hrsg.): *Springer Handbook of Speech Processing* Bd. 126. Springer, 2009. – DOI 10.1121/1.3203918. – ISBN 9783540491255, Kapitel 53, 1095–1114

[Sno55] Snow, William B.: Basic Principles of Stereophonic Sound. In: *IRE Transactions on Audio* 3 (1955), S. 42–53. http://dx.doi.org/10.1109/TAU.1955.1165407. – DOI 10.1109/TAU.1955.1165407. – ISBN 0096–1981 VO – AU–3

[SRA08] Spors, Sascha ; Rabenstein, Rudolf ; Ahrens, Jens: The Theory of Wave Field Synthesis Revisited. In: *124th AES Convention*, 2008, Convention Paper 7358

[SS06] Støfringsdal, Bård ; Svensson, Peter: Conversion of Discretely Sampled Sound Field Data to Auralization Formats. In: *J. Audio Eng. Soc* 54 (2006), Nr. 5, 380–400. http://www.aes.org/e-lib/browse.cfm?elib=13682

[TB86] Toothman, Johnson ; Brenig: *ANSI SI.II-1986: Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.* 1986

[WA01] Ward, Darren B. ; Abhayapala, Thushara D.: Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers. 9 (2001), Nr. 6, S. 697–707

[WAM09] Wu, Yan J. ; Abhayapala, Thushara D. ; Member, Senior: Theory and Design of Soundfield Reproduction Using Continuous Loudspeaker Concept. 17 (2009), Nr. 1, S. 107–116

[Wil99] Williams, Earl G.: *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* Academic Press, 1999. – ISBN 0127539603