Stefan Riedel

Compact Spherical Loudspeaker Arrays: New Ideas on Filter and Layout Design

Master's Thesis

University of Technology Graz
University of Music and Performing Arts Graz

Institute of Electronic Music and Acoustics
Head: O.Univ.-Prof. Mag.art. Dipl.-Ing. Dr.techn. Robert Höldrich

Supervisor: Dr.rer.nat. Dipl.-Ing. Franz Zotter

Graz, February 2019

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, _____ _____

Date Signature

Eidesstattliche Erklärung[1]

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am _____ _____

Datum Unterschrift

[1] Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008; Genehmigung des Senates am 1.12.2008

Abstract

Compact spherical loudspeaker arrays employ modal beam forming techniques to precisely excite wall reflections. In its performance practice it turned out that it is mostly the horizontal plane, in which walls and reflectors are close enough to create distinct auditory objects in space. Moreover, human sound localization is known to be more precise in the horizontal plane. This motivates the development of mixed-order control and dedicated mixed-order arrays that concentrate their loudspeakers in the horizontal plane. Furthermore, this thesis presents a two-band approach for the control of spherical loudspeaker arrays. In the lower band, spherical beamforming is feasible and fast crosstalk-cancellation filters can be designed. In the higher band, where beamforming is impaired due to spatial aliasing and uncontrolled interferences between the loudspeakers, an amplitude panning approach based on AllRAD is employed.

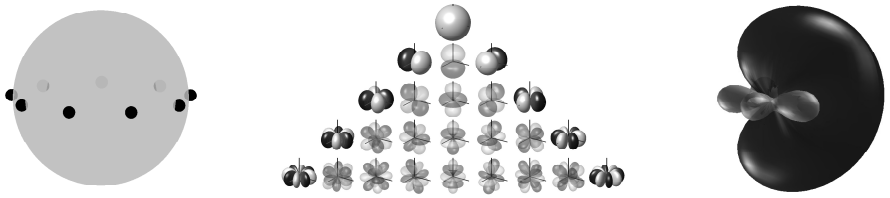
Contents

1. Introduction

In the late 1980s the idea of compact spherical loudspeaker arrays with controllable directivity was introduced by reasearchers at IRCAM and published by Warufsel *et al.* [16, 12]. Today the icosahedral loudspeaker (IKO) is used as an instrument in electro-acoustic music [21]. It is capable of creating convincing spatial trajectories along surrounding sound reflectors and is well-approved as a concert instrument.

In general, Platonic solids offer practical housings because of their symmetries and small number of facets/transducers. Classically, spherical beamforming on the dodecahedron uses 12 transducers and is limited to 2^{nd} order Ambisonic reproduction, while it is limited to 3^{rd} order with the 20 transducers on an icosahedron. Research on the topic has focused on ways to overcome the limited resolution, e.g. array-specific acoustic radiation modes have been proposed [11], but those modes would require a frequency-dependent beam encoding. Alternatively, the number of transducers per surface has been increased, e.g. to 6 on the 20 icosahedral facets [1], which, however, is only practical with high-frequency tweeters.

Figure 1.1.: The IEM icosahedral loudspeaker array in concert.



(a) (b) (c)

Figure 1.2.: Layout of a circular array with 9 transducers (a), spherical harmonics subset (b) and the beam pattern simulated with the spherical cap model at f=400 Hz, array radius $r = 0.21$ m (c).

In the current performance practice with the IKO, mostly the *horizontal* beams effectively produce auditory objects via lateral reflections from close-enough walls and baffles. Another practical aspect is the heaviness of an array with twenty broadband woofers, which typically requires two people to cooperate in the staging of the array.

Therefore, the idea of a circular horizontal transducer ring seems obvious, as it should achieve equal or better horizontal resolution, while drastically reducing the cost and weight of the array. However, the directional definition is impaired by erroneous vertical radiation lobes. In the horizontal plane, circular arrays are less robust against spatial aliasing and effectively do not reach the assumed effective order of directivity, as described in Sec. 3.3. The problem was discussed for circular microphone arrays in [9], but the performance limitation (even below spatial aliasing) is easily recognized from Fig. 1.2 (c).

A different concept is to distribute multiple (horizontal) first-order arrays such as the IEM loudspeaker cubes [5, 10], which achieves Ambisonic surround sound, but abandons the idea of a centrally staged, higher order beamformer that is capable of controlling both direction and distance of an auditory object.

The idea discussed in this thesis is to employ mixed-order Ambisonic control and array design to reduce the transducer amount while achieving a high horizontal resolution. In contrast to a single horizontal ring, transducers are added on elevated rings to achieve robustness and definition in vertical directions. Another obvious benefit over a circular array is that beams in vertical directions can be controlled by such mixed-order arrays and might be interesting in some venues.

Marschall describes mixed-order Ambisonics schemes that effectively reduce the number of microphones of an array by neglecting certain vertical spherical harmonic modes [8] in order to maintain a high horizontal resolution. As the gap between horizontal and overall resolution cannot be overly stressed for robust beamforming, alternative lattice-schemes were presented in [3]. A modified mixed-order scheme is presented here to efficiently increase the horizontal resolution of compact spherical loudspeaker arrays.

Chapter 2 of this thesis briefly discusses fundamental concepts of Fourier acoustics and defines the notation used throughout this thesis. In section 3.1 spherical beamforming is discussed for the general case of higher-order Ambisonics and for the special case of mixed-order Ambisonics. The spherical cap modal described in section 3.2 allows for the simulation of various interesting loudspeaker layouts and mixed-order control schemes.

Section 3.3 presents the proposed mixed-order schemes to increase the resolution of a dodecahedral array from 2^{nd} to 3^{rd} order, and for an icosahedral array from 3^{rd} to 4^{th} order. Additionally, new 3-ring layouts and their mixed-order schemes are described that effectively reduce the number of transducers yielding dedicated mixed-order layouts. For assessment, quality measures are discussed to evaluate the effective 2D order on the horizon in comparison to the effective overall 3D order, in order to avoid global deterioration. Both measures are based on the $\boldsymbol{r}_{\text{E}}$ vector. Chapter 5 presents a measurement-based, low-latency, two-band control model. This new model allows to minimize filter length and erroneous lobes in the high-frequency band by crossing over to AllRAD panning. Chapter 6 verifies the approach using open measurement data.

2. Fundamentals of Fourier Acoustics

2.1. The Wave Equation and its Solutions

The solutions of the wave equation are fundamental to all further work and concepts. The wave equation in the frequency domain is referred to as the Helmholtz equation

$$(\triangle + \tilde{k}^2)\, p = 0\,, \tag{2.1}$$

with the sound pressure p, the Laplace operator $\triangle$ and the wave number $\tilde{k} = \omega/c$, where ω denotes the angular frequency and c is the speed of sound. Its solution in spherical coordinates is obtained by separation of variables ($\triangle = \triangle_r + \triangle_{\varphi,\vartheta}$) and consists of a term with dependence on the radius r and a term depending on the azimuth angle φ and zenith angle ϑ [18].

Any traveling wave type solution to Eq. 2.1 can be described as an expansion of these radial and angular basis functions

$$p(\varphi, \vartheta, \tilde{k}r) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left(C_{nm}\, j_n(\tilde{k}r) + D_{nm}\, h_n^{(2)}(\tilde{k}r) \right) Y_n^m(\varphi, \vartheta)\,, \tag{2.2}$$

with the real-valued spherical harmonics $Y_n^m(\varphi, \vartheta)$ (order n and degree m)

$$Y_n^m(\varphi,\vartheta) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}\, P_n^m(\cos\,\vartheta)\cdot \begin{cases} \sqrt{2}\,\sin(|m|\varphi) & m<0\,, \\ 1 & m=0\,, \\ \sqrt{2}\,\cos(m\varphi) & m>0\,, \end{cases} \tag{2.3}$$

where zenithal dependence is defined by the associated legendre functions $P_n^m(\cos\,\vartheta)$ and azimuthal dependence by the trigonometric functions.

Physically meaningful radial solutions are the spherical Bessel function $j_n(\tilde{k}r)$ for incoming waves and the spherical Hankel function of the second kind $h_n^{(2)}(\tilde{k}r)$ for outgoing waves. The coefficients C_{nm} and D_{nm} are called the wave spectra. The spherical Bessel and spherical Hankel functions are derived from the ordinary Bessel and Hankel functions as

$$j_n(x) = \left(\frac{\pi}{2x}\right)^{1/2} J_{n+1/2}(x) = \mathfrak{R}\{h_n^{(2)}(x)\} \tag{2.4}$$

$$h_n^{(2)}(x) = \left(\frac{\pi}{2x}\right)^{1/2} H_{n+1/2}(x)\,. \tag{2.5}$$

For outgoing waves, as in the case of compact loudspeaker arrays, we speak of the exterior problem and $C_{nm} = 0$. The remaining radial function, the spherical Hankel function of the second kind $h_n^{(2)}$, is simply referred to as h_n from now on.

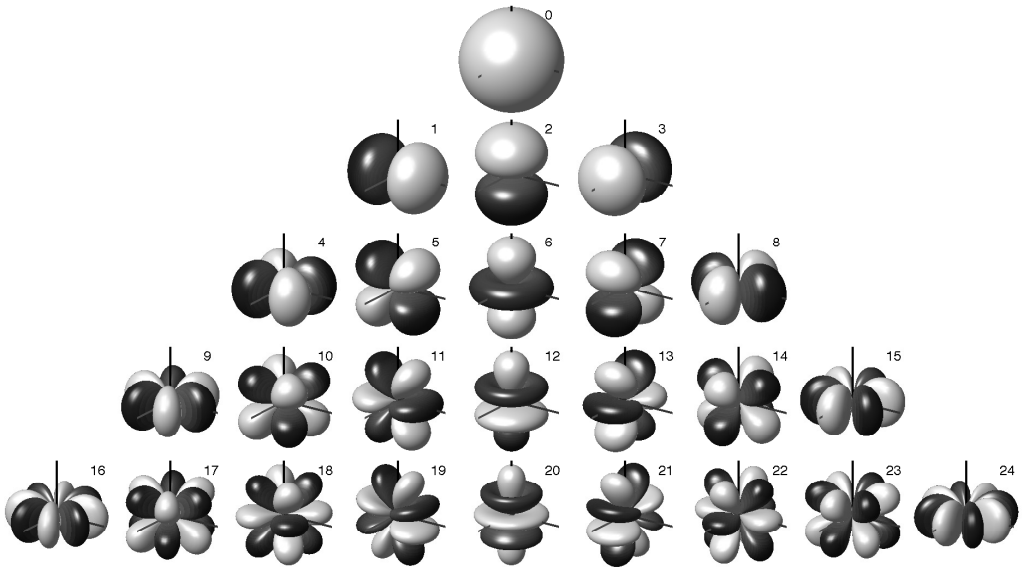


Figure 2.1.: The real-valued spherical harmonics for orders $n = 0, ..., 4$.

2.2. The Spherical Harmonics Transform

Continous Spherical Harmonics Transform

Any continous, real-valued function $p(\varphi, \vartheta)$ on the 2-sphere $\mathbb{S}^2$ can be expanded in terms of the orthonormal spherical harmonics functions

$$p(\varphi, \vartheta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \psi_{nm} \, Y_n^m(\varphi, \vartheta) \,, \tag{2.6}$$

$$\psi_{nm} = \iint_{\mathbb{S}^2} p(\varphi, \vartheta) \, Y_n^m(\varphi, \vartheta) \mathrm{d}\Omega = \mathcal{SHT}\{p(\varphi, \vartheta)\} \,, \tag{2.7}$$

with the surface element $\mathrm{d}\Omega = \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\varphi$. In case of an acoustic wave-field, $p(r_0, \varphi, \vartheta)$ is the pressure on the sphere of radius r_0 and $\psi_{nm}(r_0)$ is called the spherical wave spectrum

$$\psi_{nm}(r_0) = \iint_{\mathbb{S}^2} p(r_0, \varphi, \vartheta) \, Y_n^m(\varphi, \vartheta) \mathrm{d}\Omega. \tag{2.8}$$

Discrete Spherical Harmonics Transform

The discrete spherical harmonics transform relies on a matrix formulation of the above equations by sampling the continous functions at a limited number M of discrete points. The number and distribution of the sampling points is crucial for the inversion of the spherical harmonics matrix $\boldsymbol{Y}_\mathrm{N}$, which is necessary to calculate the coefficient vector $\boldsymbol{\psi}_\mathrm{N}$

$$\begin{bmatrix} p(\varphi_1, \vartheta_1) \\ \vdots \\ p(\varphi_\mathrm{M}, \vartheta_\mathrm{M}) \end{bmatrix} = \begin{bmatrix} Y_0^0(\varphi_1, \vartheta_1) & \dots & Y_\mathrm{N}^\mathrm{N}(\varphi_1, \vartheta_1) \\ \vdots & \vdots & \vdots \\ Y_0^0(\varphi_\mathrm{M}, \vartheta_\mathrm{M}) & \dots & Y_\mathrm{N}^\mathrm{N}(\varphi_\mathrm{M}, \vartheta_\mathrm{M}) \end{bmatrix} \begin{bmatrix} \psi_{00} \\ \vdots \\ \psi_{\mathrm{NN}} \end{bmatrix} . \tag{2.9}$$

As a requirement for a well-conditioned matrix $\boldsymbol{Y}_\mathrm{N}$ the sampling points need to be distributed appropriately (regarding the sampled set of spherical harmonics) and the number of points M is greater or equal the number of spherical harmonics

$$\mathrm{M} \geq (\mathrm{N}+1)^2 . \tag{2.10}$$

Formally, the condition number κ of a matrix is defined as the ratio between the maximum and minimum value out of the set of singular values $\{\sigma_i\}$ of the matrix and is a measure of the sensitivity of the inverted matrix ($\kappa \geq 1$).

$$\kappa(\boldsymbol{Y}_\mathrm{N}) = \frac{\sigma_\mathrm{max}(\boldsymbol{Y}_\mathrm{N})}{\sigma_\mathrm{min}(\boldsymbol{Y}_\mathrm{N})} \tag{2.11}$$

Analysis. The *left inverse* of the spherical harmonics matrix $\boldsymbol{Y}_\mathrm{N}$ is the minimum mean-square error solution (MMSE) to the overdetermined system of equations with $\mathrm{M} > (\mathrm{N}+1)^2$, where $\boldsymbol{p}$ is e.g. a pressure distribution recorded with a spherical microphone array that we want to represent in terms of spherical harmonics coefficients $\boldsymbol{\psi}_\mathrm{N}$.

We minimize the squared error between our model representation $\boldsymbol{p}_\mathrm{N}$ and the measured distribution $\boldsymbol{p}$

$$\min_{\boldsymbol{\psi_N}} ||\boldsymbol{e}||^2 \,, \qquad \text{with } \boldsymbol{e} = \boldsymbol{p}_\mathrm{N} - \boldsymbol{p} = \boldsymbol{Y}_\mathrm{N}\boldsymbol{\psi}_\mathrm{N} - \boldsymbol{p} \tag{2.12}$$

$$\frac{\partial}{\partial \boldsymbol{\psi_N}} \boldsymbol{e}^\mathrm{T}\boldsymbol{e} = 2\left(\frac{\partial}{\partial \boldsymbol{\psi_N}}\boldsymbol{e}\right)^\mathrm{T}\boldsymbol{e} = 2\,\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{e} = 2\,\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N}\boldsymbol{\psi_N} - 2\,\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{p} \overset{!}{=} 0 \tag{2.13}$$

$$\boldsymbol{\psi}_\mathrm{N} = (\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})^{-1}\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\,\boldsymbol{p} = \boldsymbol{Y}_\mathrm{N}^{+}\boldsymbol{p} = \mathcal{DSHT}\{\boldsymbol{p}\}\,. \tag{2.14}$$

The *left inverse* $(\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})^{-1}\boldsymbol{Y}_\mathrm{N}^\mathrm{T}$ inverts $\boldsymbol{p} = \boldsymbol{Y}_\mathrm{N}\boldsymbol{\psi}_\mathrm{N}$ from the left side and is referred to as *encoder* in Ambisonics.

Synthesis. If we are given a desired spectrum $\boldsymbol{\nu}_\mathrm{N}$ and need to find weights $\boldsymbol{g}$ for $\mathrm{M} > (\mathrm{N}+1)^2$ loudspeakers (underdetermined system), the optimization problem can be formulated to minimize the squared 2-norm of the weights

$$||\boldsymbol{g}||^2 \rightarrow \min\,, \tag{2.15}$$

$$\text{subject to: } \boldsymbol{\nu}_\mathrm{N} = \boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{g}\,. \tag{2.16}$$

A cost function $J(\boldsymbol{g}, \boldsymbol{\lambda})$ with the Lagrange multipliers $\boldsymbol{\lambda}$ is formulated and minimized:

$$J(\boldsymbol{g}, \boldsymbol{\lambda}) = ||\boldsymbol{g}||^2 + (\boldsymbol{\nu}_\mathrm{N}^\mathrm{T} - \boldsymbol{g}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})\,\boldsymbol{\lambda}\,, \tag{2.17}$$

$$\frac{\partial}{\partial \boldsymbol{g}^\mathrm{T}} J(\boldsymbol{g}, \boldsymbol{\lambda}) = \boldsymbol{g}_\mathrm{opt} - \boldsymbol{Y}_\mathrm{N}\,\boldsymbol{\lambda} \overset{!}{=} 0 \qquad \Rightarrow \boldsymbol{g}_\mathrm{opt} = \boldsymbol{Y}_\mathrm{N}\,\boldsymbol{\lambda}\,, \tag{2.18}$$

$$\frac{\partial}{\partial \boldsymbol{\lambda}^\mathrm{T}} J(\boldsymbol{g}, \boldsymbol{\lambda}^\mathrm{T}) = \boldsymbol{\nu}_\mathrm{N} - \boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{g} \overset{!}{=} 0 \quad \overset{\boldsymbol{g}=\boldsymbol{g}_\mathrm{opt}}{\Rightarrow} \boldsymbol{\lambda_{opt}} = (\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})^{-1}\,\boldsymbol{\nu}_\mathrm{N}\,. \tag{2.19}$$

Inserting $\boldsymbol{\lambda_{opt}}$ into $\boldsymbol{g}_\mathrm{opt}$ yields the *right inverse* $\boldsymbol{Y}_\mathrm{N}(\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})^{-1}$ of the spherical harmonics matrix, which is referred to as *decoder* in Ambisonics

$$\boldsymbol{g} = \boldsymbol{Y}_\mathrm{N}(\boldsymbol{Y}_\mathrm{N}^\mathrm{T}\boldsymbol{Y}_\mathrm{N})^{-1}\,\boldsymbol{\nu}_\mathrm{N}\,. \tag{2.20}$$

3. Array Simulation and Beam Design

3.1. Modal Beamforming and Beam Design

Directivity functions for spherical beamforming or Ambisonic panning use a finite-order, i.e. resolution-limited, representation of a Dirac delta $\delta(\Theta_\mathrm{b}^\mathrm{T}\boldsymbol{\theta} - 1)$ directed towards Θ_b and evaluated in the variable direction $\boldsymbol{\theta}$,

$$g(\boldsymbol{\theta}) = \sum_{n=0}^{\mathrm{N}} \sum_{m=-n}^{\mathrm{n}} w_{nm}\, Y_n^m(\boldsymbol{\theta}) Y_n^m(\Theta_\mathrm{b}), \tag{3.1}$$

where both of the direction unit vectors $\boldsymbol{\theta}$ and Θ_b are Cartesian unit vectors $\boldsymbol{\theta} = [\cos\varphi \sin\vartheta,\, \sin\varphi \sin\vartheta,\, \cos\vartheta]^\mathrm{T}$ depending on the azimuth angle φ and zenith angle ϑ; or φ_b and ϑ_b in case of Θ_b. Y_n^m are the spherical harmonics, and typically, to avoid side lobes, the weights w_{nm} are the max-$\boldsymbol{r}_\mathrm{E}$ weights approximated by [19]:

$$w_n = P_n\big[\cos\big(\tfrac{\pi}{180}\tfrac{137.9}{N+1.51}\big)\big], \tag{3.2}$$

Defining the vectors

$$\boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}) = [Y_n^m(\boldsymbol{\theta})]_{n=0\ldots\mathrm{N},m=-n\ldots n}, \tag{3.3}$$

$$\boldsymbol{w}_\mathrm{N} = [w_n]_{n=0\ldots\mathrm{N},m=-n\ldots n}, \tag{3.4}$$

Eq. (3.1) can be re-expressed as

$$g(\boldsymbol{\theta}) = \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta})^\mathrm{T} \mathrm{diag}\{\boldsymbol{w}_\mathrm{N}\}\, \boldsymbol{y}_\mathrm{N}(\Theta_\mathrm{b})\,, \tag{3.5}$$

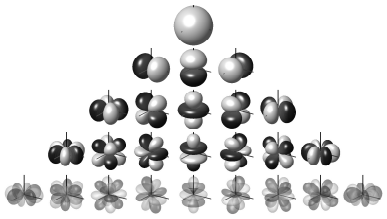
which defines a rotationally symmetric directivity pattern. The directivity function of mixed-order is different by a mask $\boldsymbol{M}$ that selects a subset of fewer

spherical harmonics, see Fig. 3.2. The mask $\boldsymbol{M}$ has $(\mathrm{N}+1)^2$ columns, each one representing a spherical harmonic, and fewer rows, of which each selects one of the harmonics to be a mixed-order component:

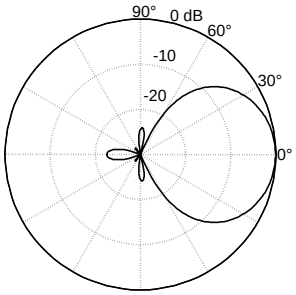
$$\begin{aligned} g(\boldsymbol{\theta}) &= \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \boldsymbol{M}^{\mathrm{T}} \mathrm{diag}\{\tilde{\boldsymbol{w}}_{\mathrm{M}}\} \boldsymbol{M}\, \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_{\mathrm{b}}) \\ &= \boldsymbol{y}_{\mathrm{M}}(\boldsymbol{\theta})^{\mathrm{T}} \mathrm{diag}\{\tilde{\boldsymbol{w}}_{\mathrm{M}}\}\, \boldsymbol{y}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{b}}). \end{aligned} \tag{3.6}$$

To get $\tilde{\boldsymbol{w}}_{\mathrm{M}} = [w_{nm}^{(\mathrm{M})}]$ we choose the vector $\boldsymbol{\theta}_{\mathrm{x}} = [1, 0, 0]^{\mathrm{T}}$ to the x direction

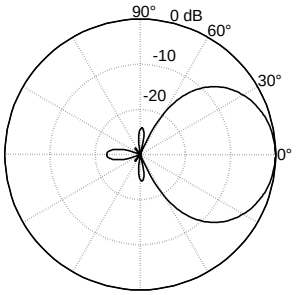
$$\tilde{w}_{nm}^{(\mathrm{M})} = w_n \frac{\sum_{n'=n}^{\mathrm{N}} [Y_{n'}^{|m|}(\boldsymbol{\theta}_{\mathrm{x}})]^2 w_{n'}}{\sum_{n'=n}^{\mathrm{N}} [Y_{n'}^{|m|}(\boldsymbol{\theta}_{\mathrm{x}})]^2 w_{n'}\, M_{n'|m|}}. \tag{3.7}$$



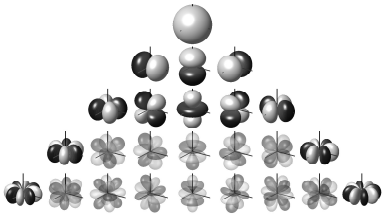
(a) Full SH-set, N = 3



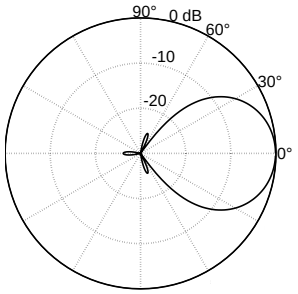
(b) Horizontal cut



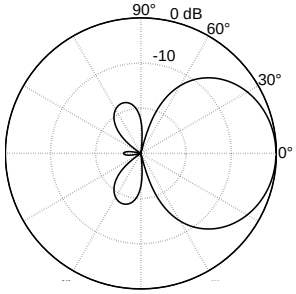
(c) Vertical cut



(d) Mixed-order subset, N = 4



(e) Horizontal cut



(f) Vertical cut

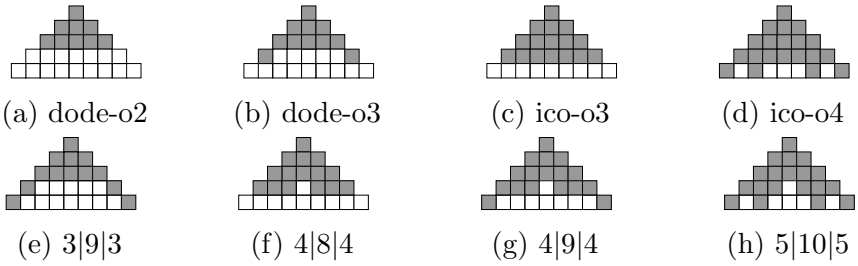
Figure 3.1.: Full SH-set of order N = 3 (a), horizontal cut (b) and vertical cut (c) through the rotationally-symmetric max-$\boldsymbol{r}_{\mathrm{E}}$-weighted pattern of a horizontal beam. Mixed-order SH-subset of order N = 4 (d), horizontal cut (e) and vertical cut (f) of the mixed-order beam pattern with the corrected weights $\tilde{\boldsymbol{w}}$. Circles indicate 0, -10, -20 dB.

Mixed-order schemes in Fig. 3.2 and associated spherical harmonic subsets can be controlled using either platonic layouts or new ring layouts consisting of a horizontal ring and two elevated rings. The platonic arrays can also be seen as 3-ring layouts, with the middle ring being a zig-zag ring of loudspeakers. That is, the dodecahedron as a $1|\tilde{10}|1$ layout and the icosahedron as a $5|\tilde{10}|5$ layout, which yields extended mixed-order control schemes for those platonic arrays. The coordinates of the new 3-ring layouts are given in Tab. 3.2.

The number of transducers n_{h} in the horizontal ring determines the maximum achievable 2D-order N_{2D} [4],

$$n_{\mathrm{h}} \geq 2\,\mathrm{N}_{\mathrm{2D}} + 1. \tag{3.8}$$

The number of transducers n_{e} in the elevated rings determines how sparse



(a) dode-o2 (b) dode-o3 (c) ico-o3 (d) ico-o4

(e) 3|9|3 (f) 4|8|4 (g) 4|9|4 (h) 5|10|5

Figure 3.2.: Control schemes for platonic and mixed-order layouts. Rows indicate the spherical harmonics order $n = 0 \dots 4$ and columns indicate the degrees $m = -n \dots n$. Scheme (b) also holds for the 373 layout.

our mixed-order control scheme is going to be. As a nomenclature we define $n_\mathrm{e}|n_\mathrm{h}|n_\mathrm{e}$ to refer to a specific layout, e.g. the 4|8|4-layout, built as pseudo-rhombicuboctahedron speaker (see Fig.4.2).

We regard the condition number κ of the mixed-order spherical harmonics matrix $\boldsymbol{Y}_\mathrm{M}$ evaluated at the transducer coordinates $\boldsymbol{\theta}_l$, to ensure a well-conditioned pseudo-inverse necessary for the array control,

$$\begin{aligned} \boldsymbol{Y}_\mathrm{M} &= \boldsymbol{M}\boldsymbol{Y}_\mathrm{N}, \\ \text{with } \boldsymbol{Y}_\mathrm{N} &= [Y_n^m(\boldsymbol{\theta}_l)]_{n=0\dots\mathrm{N},m=-n\dots n}^{l=1\dots\mathrm{L}}\ , \end{aligned} \tag{3.9}$$

Table 3.1 shows that all $\boldsymbol{Y}_\mathrm{M}$ matrices (subsets see Fig. 3.2) are sufficiently well-conditioned as $\kappa(\boldsymbol{Y}_\mathrm{M}) \ll \infty$.

	Ico.	5|10|5	4|9|4	3|9|3	4|8|4	3|7|3	Dod.
L	20	20	17	15	16	13	12
κ	2.4	1.8	1.7	1.9	1.7	2.0	1.6
$\mathrm{N_{2D}}$	4 (3)	4	4	4	3	3	3 (2)
$\mathrm{N_{3D}}$	3	≈3	≈3	2	≈3	2	2

Table 3.1.: Condition number κ of $\boldsymbol{Y}_\mathrm{M}$, number of transducers L, order of horizontal control $\mathrm{N_{2D}}$ and vertical control $\mathrm{N_{3D}}$ of simulated (mixed-order) layouts.

Lay.	4|8|4	3|9|3	3|7|3	4|9|4	5|10|5
φ-Hor.	0:45:315	0:40:320	0:51.4:308.6	0:40:320	0:36:324
φ-Up.	0:90:270	20:120:260	20:120:260	0:90:270	18:72:306
φ-Low.	45:90:315	80:120:320	80:120:320	45:90:315	54:72:342

Table 3.2.: Coordinates of mixed-order layouts. Syntax of notation is [start:step:stop] degrees of azimuthal coordinates of the horizontal, upper and lower ring of a layout. E.g. 0:40:320 denotes [0°,40°,80°,...,280°,320°]. Elevation coordinates are 0°,+45°,-45° for horizontal, upper and lower ring respectively.

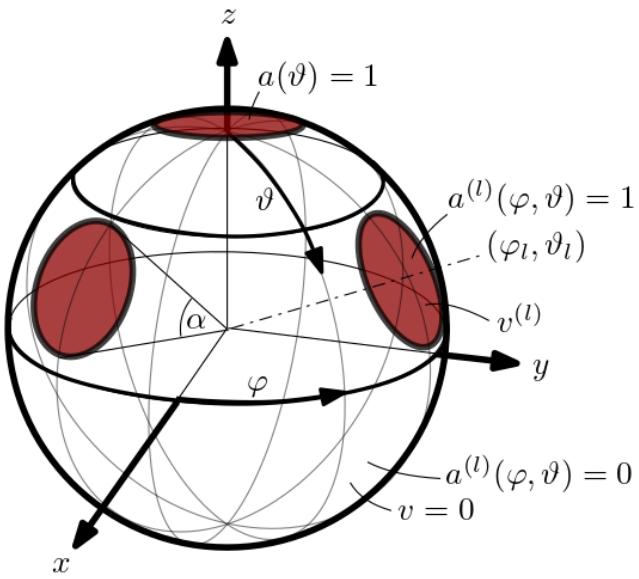


Figure 3.3.: Spherical cap model.

3.2. The Spherical Cap Model for Sound Radiation

To simulate interesting layouts acoustically, a reasonably high-order $\hat{\mathrm{N}} = 35$ model was applied that assumes moving spherical caps at the loudspeaker positions on an otherwise rigid sphere [20], see Fig. 3.3. Cap-shaped surface velocity distributions can be expressed in the spherical harmonics domain as coefficients $\nu_{nm}(\mathrm{R})$ at a radius R that can be extrapolated to a sound pressure at an analysis radius r by the following relation [18]:

$$p(\boldsymbol{\theta}, \tilde{k}r) = \mathrm{i}\rho_0 c \sum_{n=0}^{\widehat{\mathrm{N}}} \frac{h_n(\tilde{k}r)}{h_n'(\tilde{k}\mathrm{R})} \sum_{m=-n}^{n} \nu_{nm}\big|_{\mathrm{R}} Y_n^m(\boldsymbol{\theta}). \tag{3.10}$$

with the real-valued spherical harmonics Y_n^m, the spherical Hankel function of the second kind h_n and its derivative h_n', the speed of sound c, medium density ρ_0, wave number $\tilde{k} = \frac{2\pi f}{c}$ and the imaginary unit i. In our case, the velocity coefficient ν_{nm} is computed as a sum over the L cap contributions $a_{nm}^{(l)}$ weighted by the velocities $v^{(l)}$

$$\nu_{nm}\big|_{\mathrm{R}} = \sum_{l=1}^{\mathrm{L}} a_{nm}^{(l)}\, v^{(l)}. \tag{3.11}$$

An aperture function $a^{(l)}(\varphi, \vartheta)$ that equals one in the area of the l-th cap (opening angle α, direction $\boldsymbol{\Theta}_l$) and zero on the rigid sphere is defined by.

$$a^{(l)}(\varphi, \vartheta) = u\left(\boldsymbol{\theta}^T \boldsymbol{\Theta}_l - \cos\frac{\alpha}{2}\right), \tag{3.12}$$

$$\boldsymbol{\theta} = \begin{pmatrix} \cos(\varphi)\sin(\vartheta) \\ \sin(\varphi)\sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix}, \qquad \boldsymbol{\Theta}_l = \begin{pmatrix} \cos(\varphi_l)\sin(\vartheta_l) \\ \sin(\varphi_l)\sin(\vartheta_l) \\ \cos(\vartheta_l) \end{pmatrix}, \tag{3.13}$$

where the unit step function $u(.)$ equals one for positive arguments, zero otherwise.

To get the harmonic coefficients $a_{nm}^{(l)}$ of a specific piston at $\boldsymbol{\Theta}_l$ we compute the integral for an axisymmetric piston location and exploit the shift property of the $\mathcal{SHT}$. Thus, the $\mathcal{SHT}$ of equation (3.12) yields

$$a_{nm}^{(l)} = \mathcal{SHT}\{a^{(l)}(\varphi, \vartheta)\} = Y_n^m(\boldsymbol{\Theta}_l) \cdot \mathcal{SHT}\{a(\vartheta)\} \tag{3.14}$$

$$= Y_n^m(\boldsymbol{\Theta}_l) \cdot \frac{2n+1}{2} \int_0^{\alpha/2} P_n(\cos\ \vartheta) \sin\vartheta d\vartheta \tag{3.15}$$

$$= Y_n^m(\boldsymbol{\Theta}_l) \cdot a_n. \tag{3.16}$$

The integral is computed by using a recurrence formula for the Legendre polynomials and yields [18]

$$a_n = \begin{cases} \frac{1}{2}\left(P_{n-1}(\cos\frac{\alpha}{2}) - P_{n+1}(\cos\frac{\alpha}{2})\right) & n \geq 1 \\ \frac{1}{2}(1-\cos\frac{\alpha}{2}) & n = 0 \end{cases}. \tag{3.17}$$

The model can be written in matrix form

$$p(\boldsymbol{\theta}) = \boldsymbol{y}(\boldsymbol{\theta})^{\mathrm{T}}\mathrm{diag}\{\boldsymbol{h}\}\mathrm{diag}\{\boldsymbol{a}\}\boldsymbol{Y}\,\boldsymbol{v} = \boldsymbol{y}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{Q}\,\boldsymbol{v}. \tag{3.18}$$

The matrices and vectors used are defined as

$$\boldsymbol{h} = \left[i\rho_0 c\frac{h_n(\tilde{k}r)}{h_n'(\tilde{k}\mathrm{R})}\right]_{n=0\ldots\widehat{\mathrm{N}},m=-n\ldots n} \tag{3.19}$$

$$\boldsymbol{a} = [a_n]_{n=0\ldots\widehat{\mathrm{N}},m=-n\ldots n} \tag{3.20}$$

$$\boldsymbol{y}(\boldsymbol{\theta}) = [Y_n^m(\boldsymbol{\theta})]_{n=0\ldots\widehat{\mathrm{N}},m=-n\ldots n} \tag{3.21}$$

$$\boldsymbol{Y} = [\boldsymbol{y}(\boldsymbol{\theta}_1)\ldots\boldsymbol{y}(\boldsymbol{\theta}_{\mathrm{L}})]. \tag{3.22}$$

Now we introduce a control matrix $\boldsymbol{C}$ that yields the cap velocities $\boldsymbol{v}$ for a desired beam pattern in the controllable mixed-order subscpace, encoded as $\mathrm{diag}\{\tilde{\boldsymbol{w}}_{\mathrm{M}}\}\boldsymbol{y}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{b}})$,

$$p_{\mathrm{M}}(\boldsymbol{\theta}) = \boldsymbol{y}_{\mathrm{M}}(\boldsymbol{\theta})^{\mathrm{T}}\,\boldsymbol{Q}_{\mathrm{M}}\,\boldsymbol{C}\,\mathrm{diag}\{\tilde{\boldsymbol{w}}_{\mathrm{M}}\}\,\boldsymbol{y}_{\mathrm{M}}(\boldsymbol{\theta}_{\mathrm{b}}). \tag{3.23}$$

To match the above with Eq. (3.7) the control matrix $\boldsymbol{C}$ becomes using the pseudo inverse $(.)^{+}$,

$$\boldsymbol{Q}_{\mathrm{M}}\boldsymbol{C} = \boldsymbol{I} \Longrightarrow \boldsymbol{C} = \boldsymbol{Q}_{\mathrm{M}}^{+}. \tag{3.24}$$

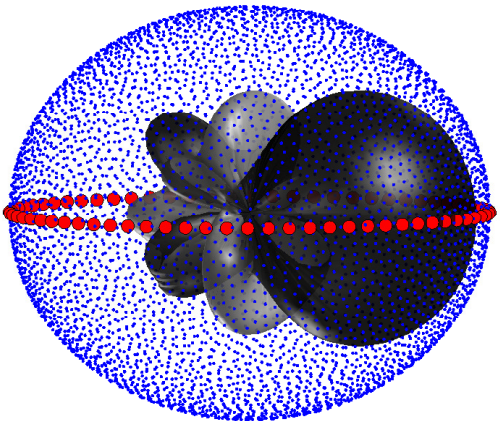


Figure 3.4.: Blue dots indicate the $J = 5100$ point t-design sampling to calculate a 3D effective order $N_{\text{eff,3D}}$ for a simulated beam pattern. Red markers indicate $J = 72$ equi-angular sampling points to calculate a 2D effective order $N_{\text{eff,2D}}$.

3.3. Simulation of Loudspeaker Layouts

A simulation study was conducted calculating effective orders of horizontal beams for the various layouts. Therefore, the simulated beam pattern (Eq. (3.18)) was regularly sampled with a J = 5100 point t-design [6] to calculate the energy vector $\boldsymbol{r}_{\text{E}}$ as a measure of directivity

$$\boldsymbol{r}_{\text{E}} = \frac{\sum_{j=1}^{\text{J}} \boldsymbol{\Theta}_j \, p(\boldsymbol{\Theta}_j)^2}{\sum_{j=1}^{\text{J}} p(\boldsymbol{\Theta}_j)^2} \; . \tag{3.25}$$

The effective order $N_{\text{eff,3D}}$ is calculated from the length $\|\boldsymbol{r}_{\text{E}}\|$ of the energy vector [4, 19]

$$N_{\text{eff,3D}} = \frac{\pi}{180} \frac{137.9}{\arccos \|\boldsymbol{r}_{\text{E}}\|} - 1.51 \; . \tag{3.26}$$

An effective order $N_{\text{eff,2D}}$ is calculated by a circular, even-angular sampling (J = 72, i.e. 5° steps) of the beam pattern in the horizontal plane

$$N_{\text{eff,2D}} = \frac{\pi}{180} \frac{90}{\arccos \|\boldsymbol{r}_{\text{E}}\|} - 1 \; . \tag{3.27}$$

3.3.1. Low-Frequency Simulation: Modal Beamforming

The results shown in Fig. 3.5 indicate that Platonic arrays can take an order jump in the 2D metric, although there is a slight loss in the 3D metric for the dodecahedral layout. The simulations also show that mixed-order layouts like the 4|8|4-layout and the 3|9|3-layout achieve equal 2D ratings compared to the icosahedral layout, while saving 4 to 5 loudspeakers.

Figs. 3.6 to 3.8 visualize the degrading effect of aliasing for a selection of layouts. Beam patterns are depicted at a frequency of 200 Hz and at a higher frequency of 800 Hz. Note that the 4|8|4-layout and the icosahedral layout create identical horizontal beam patterns at low frequencies.

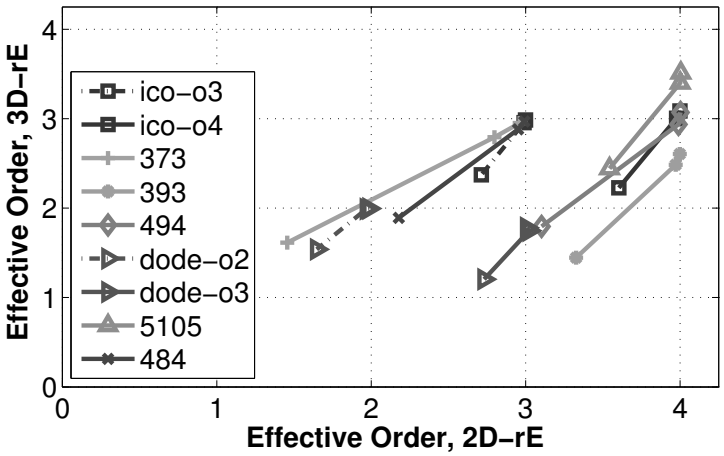
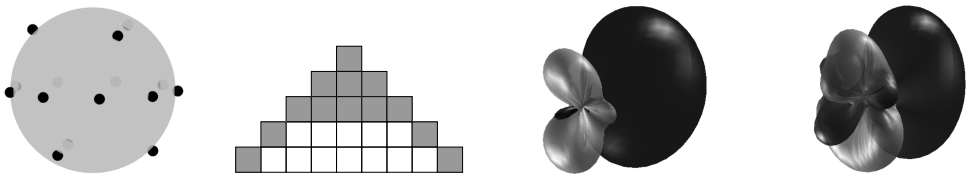
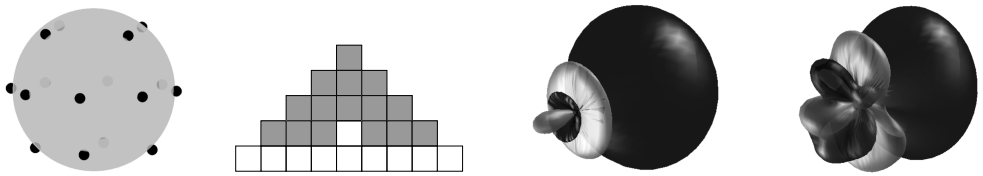


Figure 3.5.: 2D/3D effective orders of horizontal beam. Markers indicate $[200, 400, 800]$ Hz, descending from north-east to south-west due to spatial aliasing. Platonic arrays gain a full order (in 2D) with mixed-order control. The 393-array achieves 4th order in 2D rating with 5 speakers less than an icosahedral array. Longer trajectories indicate less robustness against spatial aliasing.



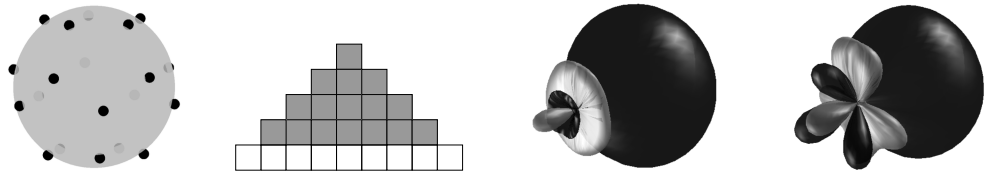
(a) (b) (c) (d)

Figure 3.6.: Layout of the 3|9|3-array (a), spherical harmonics subset (b) and mixed-order beam pattern simulated with the spherical cap model at f=200 Hz (c) and f=800 Hz (d).



(a) (b) (c) (d)

Figure 3.7.: Layout of the 4|8|4-array (a), spherical harmonics subset (b) and mixed-order beam pattern simulated with the spherical cap model at f=200 Hz (c) and f=800 Hz (d).



(a) (b) (c) (d)

Figure 3.8.: Layout of the icosahedral array (a), spherical harmonics subset (b) and 3rd order beam pattern simulated with the spherical cap model at f=200 Hz (c) and f=800 Hz (d).

3.3.2. High-Frequency Simulation: AllRAD-based Panning

So far simulations were restricted to the frequency range in which modal beamforming is feasible. For high frequencies spatial aliasing heavily impairs the directivity and causes erroneous side-lobes. An AllRAD-based panning approach is simulated and compared with the classical beamforming approach to investigate possible improvements using the loudspeaker's naturally focused radiation at high frequencies.

By encoding the beam direction $\boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_\mathrm{b})$ with a sufficiently high order, e.g. $\mathrm{N} = 5$ or higher, Allround-Ambisonic Decoding (AllRAD) [19] approximates VBAP [14], while remaining within the ambisonic domain. First, we decode to a virtual layout of $\mathrm{J} = 5100$ t-design points and therefore evaluate the spherical harmonics up to order N at those points.

$$\boldsymbol{Y}_\mathrm{N,J} = [\boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_\mathrm{J})]^\mathrm{T}. \tag{3.28}$$

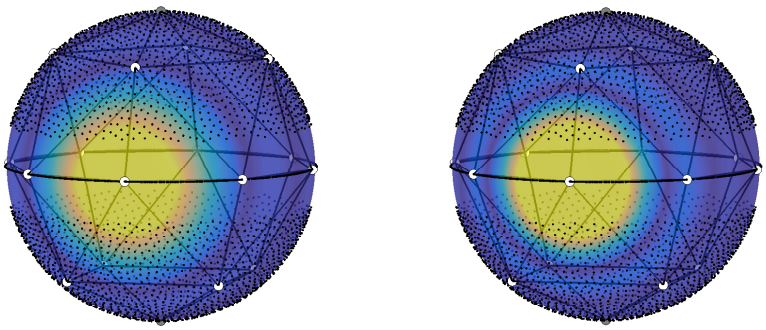
The $\mathrm{L} \times \mathrm{J}$ matrix $\boldsymbol{G}$ renders the $\mathrm{J} = 5100$ virtual sources onto L loudspeakers

$$\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_\mathrm{J}]. \tag{3.29}$$

Note that only 1, 2 or in most cases 3 values of $\boldsymbol{g}_j$ are non-zero, depending on the direction $\boldsymbol{\theta}_j$, with $j = 1, \ldots, \mathrm{J}$. As a source beam we choose a max-$\boldsymbol{r}_\mathrm{E}$ weighted 5^th or 7^th order beam, see Fig. 3.9, and therefore need to apply order dependent weights $\boldsymbol{w}_\mathrm{N}$. We arrive at the $\mathrm{L} \times (\mathrm{N}+1)^2$ panning decoder matrix $\boldsymbol{D}$

$$\boldsymbol{D} = \frac{4\pi}{\mathrm{J}}\, \boldsymbol{G}\, \boldsymbol{Y}_\mathrm{N,J}\, \mathrm{diag}\{\boldsymbol{w}_\mathrm{N}\}. \tag{3.30}$$

By pulling all t-design points that are within an elevation angle of e.g. $\beta = \pm 15°$ onto the horizon, we can avoid energy being distributed up- or downwards for horizontal panning directions. In particular, for the icosahedral array we enforce pairwise panning along the the zig-zag of horizontal loudspeakers similar to a true ring layout as seen in Fig. 3.10(c) by creating a virtual ring of loudspeakers that is decoded to the actual zig-zag layout. Additionally, layouts may require imaginary loudspeakers to enable a symmetric triangulation. A valid triangulation can be achieved by finding a convex hull to the set of points, that means finding the smallest convex set that contains the points [2]. For

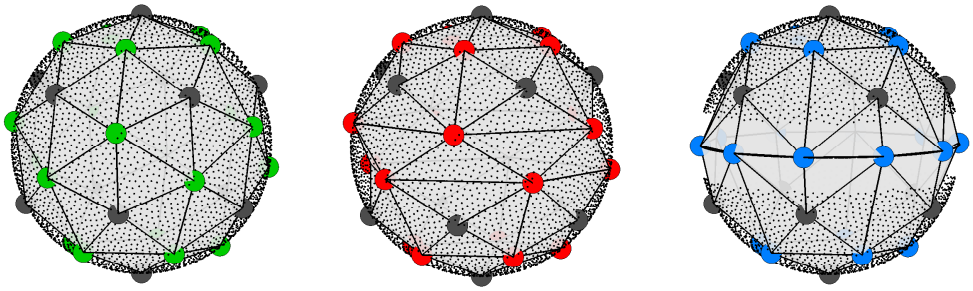


(a) Panning-Beam N = 5 (b) Panning-Beam N = 7

Figure 3.9.: Max-$\boldsymbol{r}_{\mathrm{E}}$ beams of order N = 5 or N = 7 are sufficiently narrow to approximate panning via AllRAD, as the magnitude of the beam pattern shows a quick decay (yellow to blue) around the desired panning direction. Exemplary for the 4|8|4-layout.

the icosahedral array, triangulations that rely on such search algorithms are depicted in Fig. 3.10 (a) and (c). However, we might also be interested in non-convex triangulations such as the one depicted in Fig. 3.10 (b).

A comparison of those triangulations based on the effective order $N_{\mathrm{2D,eff}}$ of a horizontal panning direction has been conducted. The results in Fig. 3.11 indicate that a virtual ring with a modified t-design is the best solution. Although the spherical cap modal is too idealistic in the high-frequency range and absolute values might differ in reality, the order of performance is still valid and confirmed by measurements. Especially the model-based comparison of different arrays shown in Fig. 3.12 and Fig. 3.13 is reflected by the measurement results in chapter 6. The 4|8|4-array exhibits the highest focus for high frequencies, as it happens that beam directions are on-axis loudspeaker directions. In between two horizontal speakers the order is likely to drop down to values of an icosahedral array.



(a) Equi-Triangular (b) Zig-Zag Triangles (c) Virtual Ring

Figure 3.10.: For the icosahedral array, different triangulations have been investigated. Big colored markers are real loudspeakers, big grey markers imaginary loudspeakers to be downmixed. Small black dots indicate the J = 5100 t-design points. In case of the virtual ring (c), additionally the t-design points close to the horizon are pulled towards the horizon, to decrease the spread of energy.

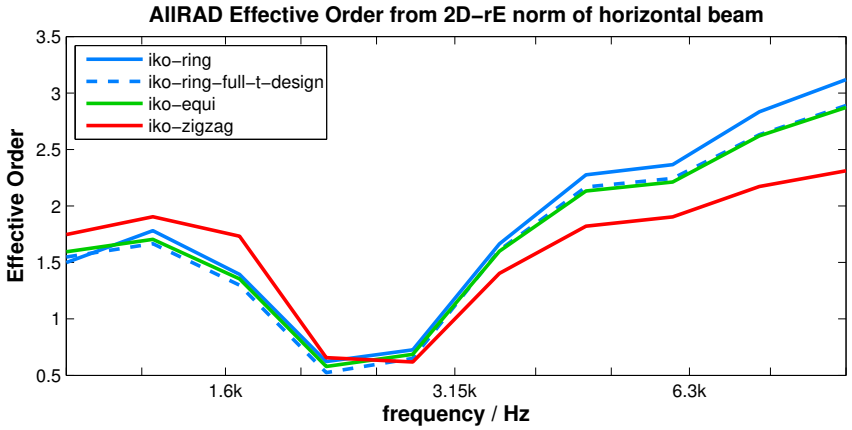


Figure 3.11.: Simulated 2D effective order of different triangulations for the icosahedral arary.

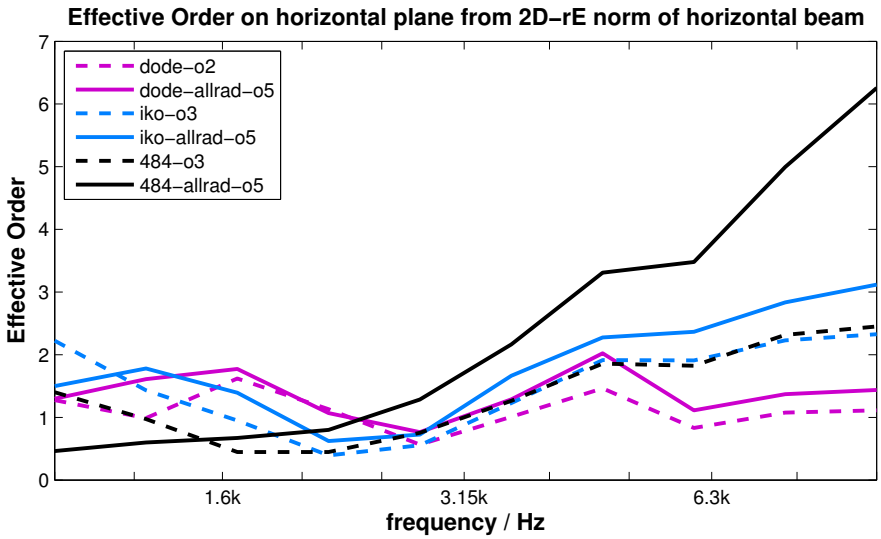


Figure 3.12.: High-frequency simulation comparing the 2D effective order of modal beamforming and AllRAD-based panning. The crossover for arrays simulated with a radius $r = 0.21$m would be around 1 to 1.5 kHz.

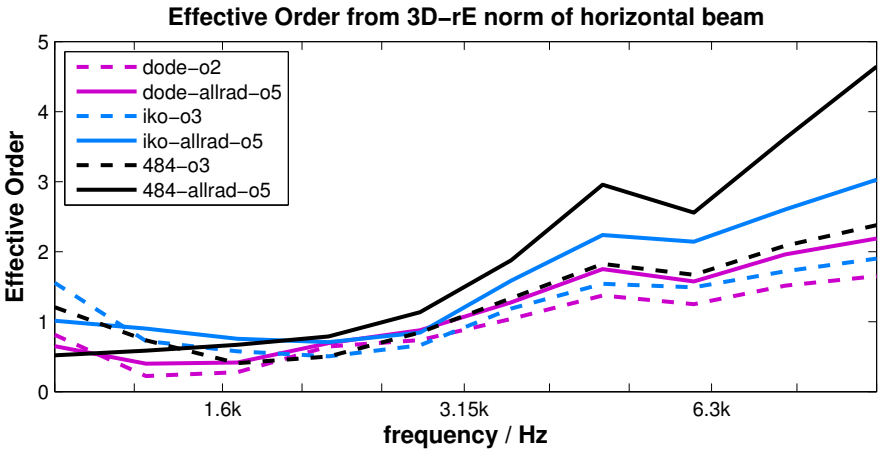


Figure 3.13.: High-frequency simulation comparing the 3D effective order of modal beamforming and AllRAD-based panning. The crossover for arrays simulated with a radius $r = 0.21$m would be around 1 to 1.5 kHz.

4. Array Design and Measurements

4.1. Design of the $4|8|4$-Array

We decided to build a wooden prototype of the $4|8|4$-layout ($r = 0.18$ m, 5" Satori MW13P-4 transducers from *SB Acoustics*) as it can be built as a regular polyhedron composed of equi-sided square and triangular surfaces (a pseudo-rhombicuboctahedron, see Fig.4.1). Additionally, the number of required D/A-converter channels to control the array is 16, which is very practical.

The series of pictures in Fig. 4.2 shows the process of glueing together the squares and triangles. The loudspeakers are mounted into the 16 squares and four of the triangles are used to place 4-channel NL8MPR connectors. The bottom square is mounted with a K&M 19656 loudspeaker flange.

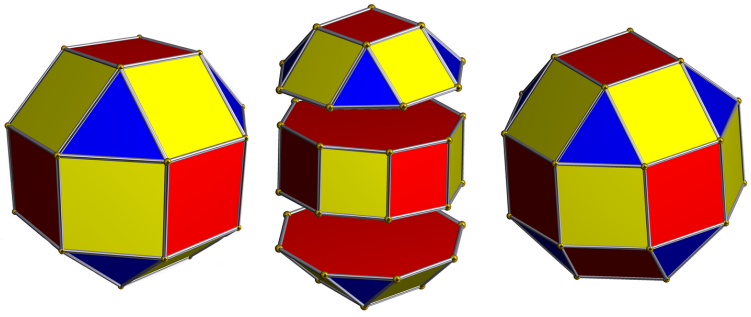


Figure 4.1.: Left: The Rhombicuboctahedron Middle: Exploded Rhombicuboctahedron Right: Pseudo-rhombicuboctahedron by rotation of the bottom part. Credit: [17]

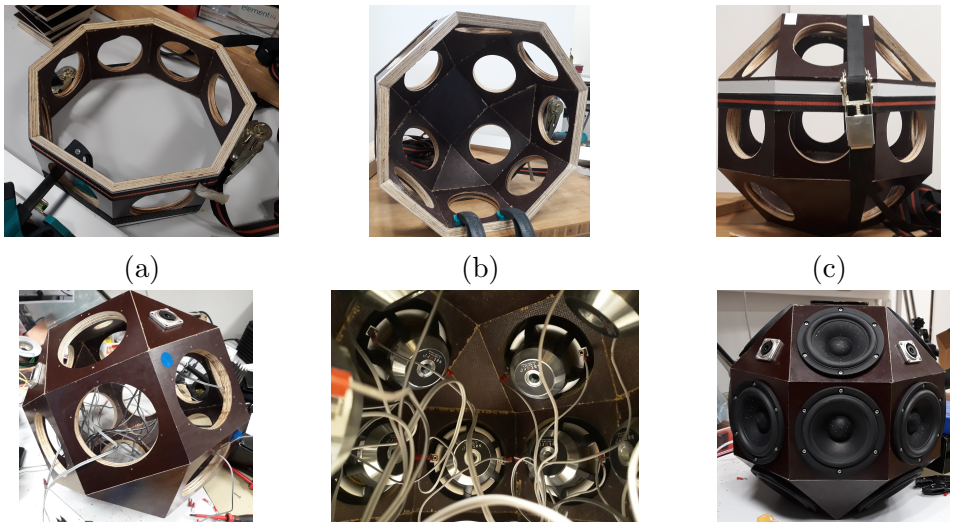


Figure 4.2.: The building process of the pseudo-rhombicuboctahedral array.

4.2. Design of the $3|9|3$-Array

Furthermore, an OpenSCAD model of a 3|9|3-array housing was created, see Fig. 4.3, which is freely downloadable[1]. The housing has been 3D-printed with a radius of $r = 0.12$ m and is mounted with 2.5" transducers from *SB Acoustics* (SB65WBAC25-4). The 393-array only counts 15 loudspeakers (smaller and cheaper transducers) which leads to the idea of staging the array with an additional subwoofer and yields a new "$15 + 1$" concept. The connectors and the bottom flange are the same as for the 4|8|4-array.

[1] `https://github.com/stefanriedel/the-393-array`

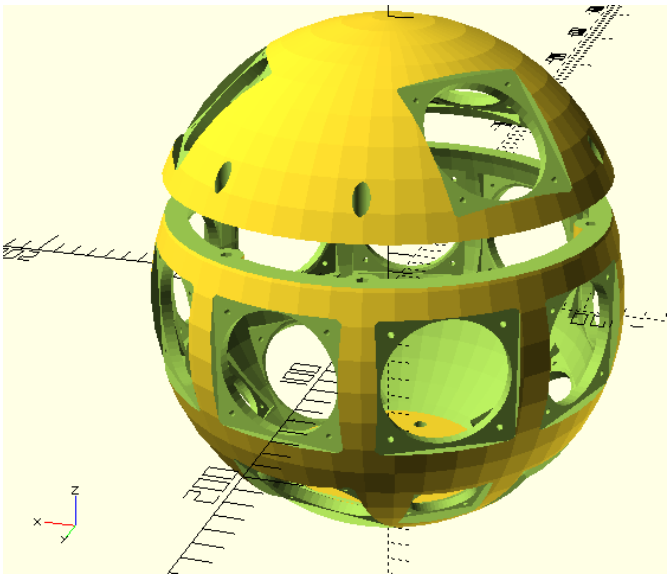


Figure 4.3.: OpenSCAD model of the 3|9|3 housing. Top and bottom part are 3D-printed and screwed together.

4.3. MIMO voltage-to-velocity measurements

To measure the crosstalk between the loudspeakers that are mounted in a common housing, voltage-to-velocity measurements are conducted by means of Laser-Doppler-Vibrometry (see Fig. 4.4). Later in Sec. 5.4 a crosstalk-cancellation filter is designed based on the $L \times L$ MIMO system of measured frequency responses.

Velocity measurements are free from environmental noise and valid for low- to mid-range frequencies until the modal breakup of a loudspeaker cone. As we are measuring at a single, central point of the transducer the measurement does not properly characterize the behaviour at high frequencies. The measurement data is freely available for download[2].

[2] `http://phaidra.kug.ac.at/o:77571`, `http://phaidra.kug.ac.at/o:77568`

(a) (b)

Figure 4.4.: MIMO voltage-to-velocity measurements with a laser-doppler vibrometer aiming at the center of a loudspeaker cone. All loudspeakers are excited with an exponentially swept-sine and the crosstalk is measured at the loudspeaker the LDV is aimed at.

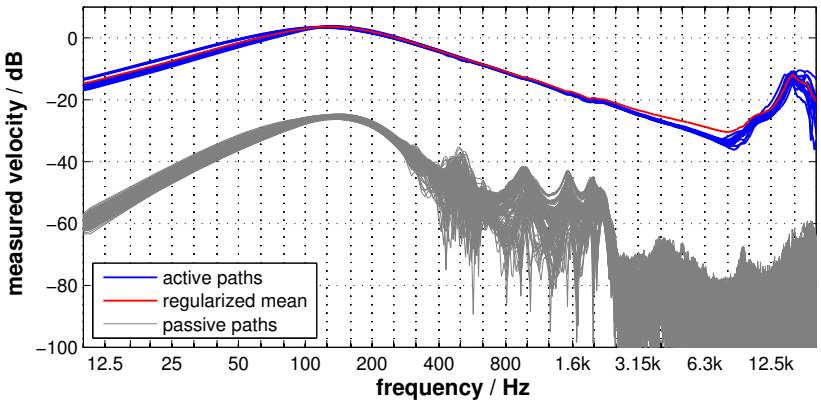
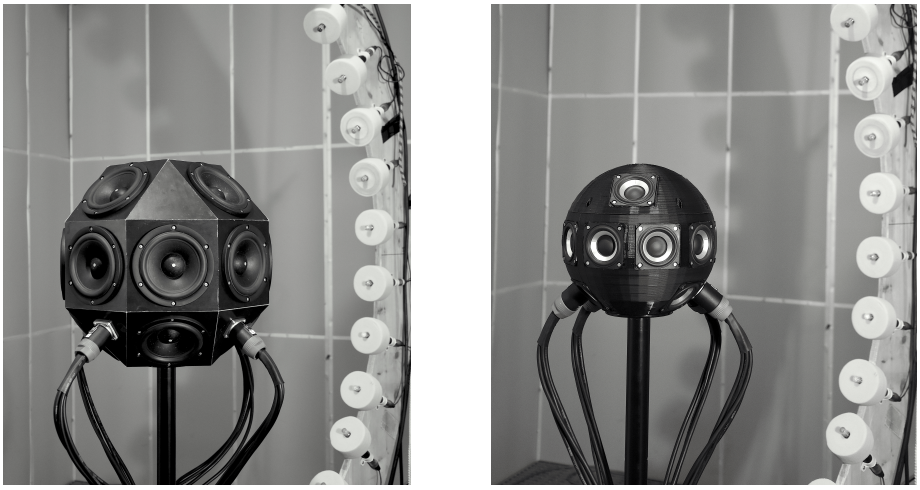


Figure 4.5.: Loudspeaker velocity measurements of the 3|9|3-array. Active paths are shown in blue, their regularized mean in red and passive paths (due to acoustic crosstalk) are shown in grey.

4.4. MIMO voltage-to-sound pressure measurements

Directional impulse response (DIR) measurements from all of the L loudspeakers to a surrounding spherical microphone array with an effective number of $R = 648$ measurement points are conducted. The measurements are used to later evaluate the directivity of the arrays and also for an overall timbral equalization (Sec. 5.6).

The loudspeaker array is positioned on a remotely-controllable turntable and surrounded by a semi-circular microphone array of 18 microphones, see Fig. 4.6. By rotating the loudspeaker array in 10° steps a measurement resolution of $10° \times 10°$ (azimuth $\times$ zenith) is achieved, see Fig. 4.7. The result is a set of $L \times 648$ impulse responses. The measurement data is available in the Spatially Oriented Format for Acoustics (SOFA) and can be downloaded [3].



(a) (b)

Figure 4.6.: MIMO sound pressure measurements with a surrounding microphone array.

[3] `http://phaidra.kug.ac.at/o:77431`, `http://phaidra.kug.ac.at/o:77567`

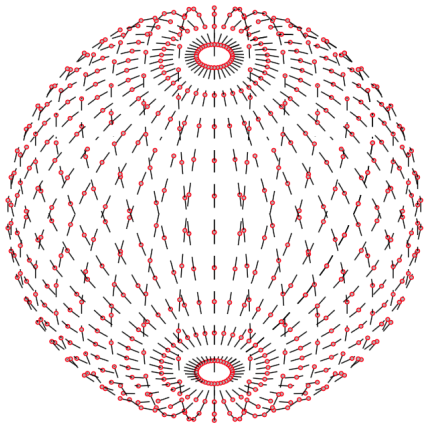


Figure 4.7.: Sampling layout of the MIMO sound pressure measurements (taken from [13]). The figure shows the $R = 36 \cdot 18 = 648$ sampling points, resulting in a $10° \times 10°$ (azimuth and zenith) resolution.

5. Filter Design for Variable-Directivity Control

5.1. System Overview and Two-Band Approach

As shown in the control overview, figure 5.1, a two-band approach is proposed. Since we're sampling the sphere at discrete points, we're only able to control the soundfield up to a certain frequency, the spatial aliasing frequency. From there on the beamforming will be heavily impaired. Around this spatial aliasing frequency, we set the crossover to a panning approach. To stay within the ambisonic workflow and framework, an AllRAD [19] based approach is presented. By encoding sufficiently high (e.g. 5th or 7th order), we can use this relatively narrow encoding beam as a panning direction.

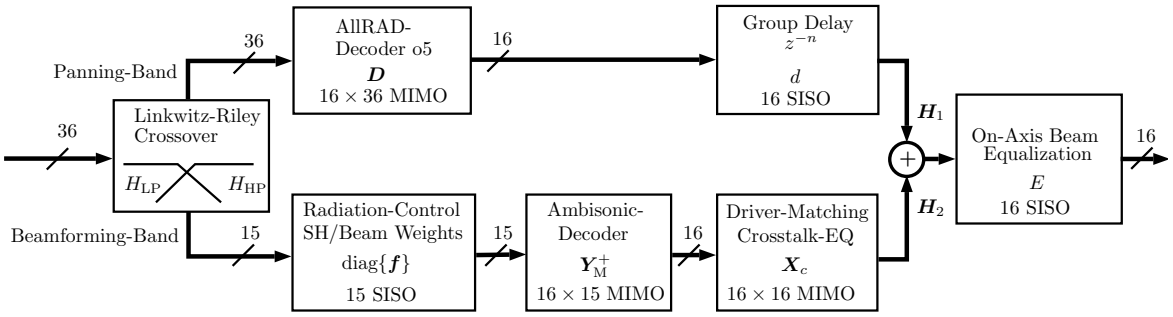


Figure 5.1.: Two-band control system diagram.

5.2. Linkwitz-Riley Band Splitting

A Linkwitz-Riley crossover is composed of a parallel high-pass and low-pass cascade of Butterworth filters. A single butterworth filter shows a -3 dB drop at the cut-off frequency f_c, a cascade of two thus exhibits -6 dB which yields a flat response when adding up the two bands of the crossover. For our filter design we used a Linkwitz-Riley filter of 6th order, cascading two Butterworth filters of 3rd order. This is useful to get a steep crossover between the beamforming-band and the panning-band.

A thorough definition of the Butterworth filters is useful as the regularization filterbank in subsec. 5.3.2 is also designed from Butterworth filters. The transfer function for an n^{th} order Butterworth lowpass filter is given as

$$L_n\left(s_c\right) = \frac{1}{D_n\left(s_c\right)}\,, \tag{5.1}$$

with the denominator $D_n(s)$ being a Butterworth polynomial with the normalized frequency $s_c = \mathrm{i}\frac{\omega}{\omega_c}$

$$\begin{aligned} D_n\left(s_c\right) &= \prod_{k=1}^{\frac{n}{2}} \left[s_c^2 + 2\, s_c \cos\left(\frac{2k-1}{2n}\pi\right) + 1 \right] && n \;= \text{even}\,, \\ D_n\left(s_c\right) &= \prod_{k=1}^{\frac{n-1}{2}} \left[s_c^2 + 2\, s_c \cos\left(\frac{2k-1}{2n}\pi\right) + 1 \right] (s_c + 1) && n = \text{odd}\,. \end{aligned} \tag{5.2}$$

A butterworth highpass filter is given by lowpass-highpass transformation replacing s_c with $1/s_c$ in the above equations:

$$H_n(s_c) = L_n(1/s_c)\,. \tag{5.3}$$

For example, a second-order highpass filter is derived to

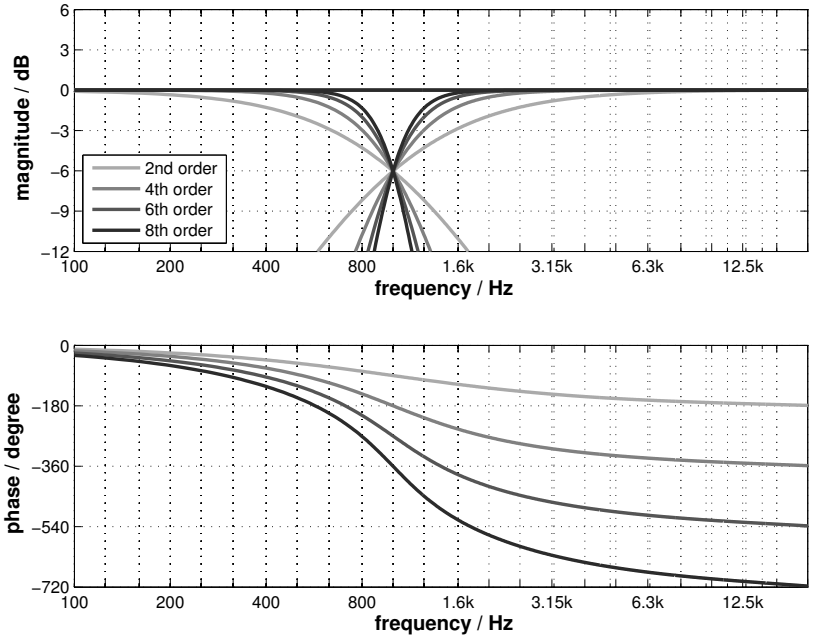


Figure 5.2.: Linkwitz-Riley Crossover of different orders at $f_c = 1$ kHz.

$$H_2(s_c) = L_2(1/s_c) = \frac{1}{(1/s_c)^2 + \sqrt{2}\,(1/s_c)\, + 1} = \frac{s_c^2}{s_c^2 + \sqrt{2}\,s_c\, + 1}\,. \tag{5.4}$$

A single Linkwitz-Riley crossover of $(2n)^{\text{th}}$ order $C_{2n}(s_c)$ is then defined as a parallel section of cascaded Butterworth filters

$$C_{2n}(s_c) = L_n(s_c)^2 + (-1)^n H_n(s_c)^2\,. \tag{5.5}$$

5.3. Radial and Angular Control for the Beamforming Band

In our actual filter design we apply a regularization filterbank diag$\{\boldsymbol{r}\}$ to the control matrix $\boldsymbol{C}$ in Eq. (3.24) to limit the loudspeaker excursion. Moreover, we can separate the matrix into a well-conditioned frequency-independent decoder $\boldsymbol{Y}_{\mathrm{M}}^{+}$, see Tab. 3.1, and a diagonal component diag$\{\boldsymbol{f}\}$ that contains the regularized inverse of diag$\{\boldsymbol{h}_{\mathrm{M}}\}$, the inverse of diag$\{\boldsymbol{a}_{\mathrm{M}}\}$ as well as the weights $\tilde{\boldsymbol{w}}_{\mathrm{M}}$

$$\begin{aligned}\tilde{\boldsymbol{C}} &= \left(\operatorname{diag}\{\boldsymbol{h}_{\mathrm{M}}\}\operatorname{diag}\{\boldsymbol{a}_{\mathrm{M}}\}\boldsymbol{Y}_{\mathrm{M}}\right)^{+}\operatorname{diag}\{\tilde{\boldsymbol{w}}_{\mathrm{M}}\}\operatorname{diag}\{\boldsymbol{r}\} \\ &= \boldsymbol{Y}_{\mathrm{M}}^{+}\operatorname{diag}\{\boldsymbol{f}\} = \boldsymbol{Y}_{\mathrm{M}}^{\mathrm{T}}(\boldsymbol{Y}_{\mathrm{M}}\boldsymbol{Y}_{\mathrm{M}}^{\mathrm{T}})^{-1}\operatorname{diag}\{\boldsymbol{f}\}. \end{aligned} \tag{5.6}$$

The regularization is based on the limited-excursion design presented in [21] and employs a low-latency filterbank described in [7].

5.3.1. Angular Control

The frequency independent $L \times n_{\mathrm{M}}$ decoder matrix $\boldsymbol{Y}_{\mathrm{M}}^{+}$ computes L velocities from an input pattern composed of n_{M} mixed-order harmonics

$$\boldsymbol{Y}_{\mathrm{M}}^{+} = \boldsymbol{Y}_{\mathrm{M}}^{\mathrm{T}}(\boldsymbol{Y}_{\mathrm{M}}\boldsymbol{Y}_{\mathrm{M}}^{\mathrm{T}})^{-1}\,. \tag{5.7}$$

5.3.2. Radial Control

A direct inversion of the radial component diag$\{\boldsymbol{h}_{\mathrm{M}}\}$ leads to instable filters, see fig. 5.3. We need to apply a filterbank diag$\{\boldsymbol{r}\}$ that stabilizes the radial filters to achieve a limited loudspeaker excursion, while maintaining constant loudness throughout the frequency range.

Unlimited radial filters exhibit slopes of $1/f^{n+1}$. To limit the excursion instead of velocity $|x_l| < x_{\max}$, we need an additional factor $1/f$, as the excursion is defined as the integral of the velocity. This yields high-pass limitation filters

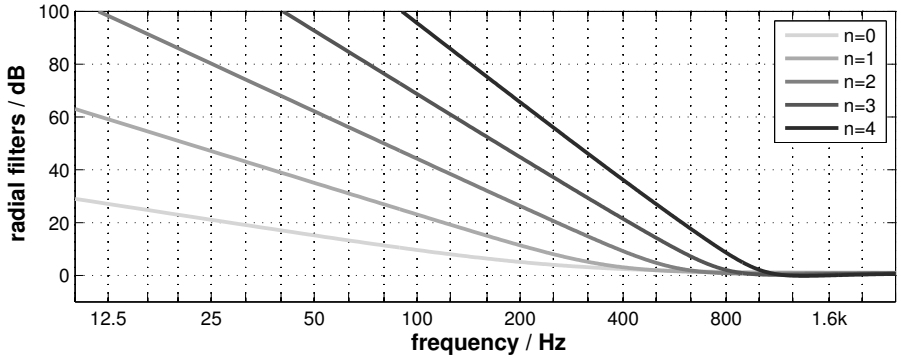


Figure 5.3.: Radial filters up to 4th order for $r_v = 0.21$m.

with slopes of f^{n+2}, but if we want to only drive efficiently radiated modes, it is necessary to use slopes of at least f^{n+3}.

A low-latency implementation is built of cascaded (second-order section) butterworth filters (Linkwitz-Riley, see. Sec. 5.2). For this reason, we can only design even order slopes.

Using the definition of Butterworth filters with the normalized frequency $s_c = \frac{s}{\omega_c} = \mathrm{i}\frac{\omega}{\omega_c}$ and the cut-off frequencies w_c for $c = 0, ..., 4$ we get a filterbank of Linkwitz-Riley bandpass crossovers

$$B_0(s) = H_2(s_0)^2\, L_2(s_1)^2\,, \tag{5.8}$$
$$B_1(s) = H_2(s_1)^2\, L_3(s_2)^2\,, \tag{5.9}$$
$$B_2(s) = H_3(s_2)^2\, L_3(s_3)^2\,, \tag{5.10}$$
$$B_3(s) = H_3(s_3)^2\, L_4(s_4)^2\,, \tag{5.11}$$
$$B_4(s) = H_4(s_4)^2\,. \tag{5.12}$$

Figure. 5.2 depicts the phase responses of even-order Linkwitz-Riley crossovers, as we find them in our regularization filterbank. Since bands of different cut-off frequencies and slopes exhibit different (shifted) phase responses, the addition of those bands leads to notches in the sum response at the frequency where the phase differs by 180 degrees.

A solution is to ensure that all bands have the same phase response by applying allpass-filters that correct the phase of every bandpass to a common phase

response. So far, the phase of $B_j(s)$ is composed of two slopes, except $B_4(s)$ which is just a highpass and has just one slope.

We are adding to each of these bandpass filters the phase of the other Linkwitz-Riley crossovers. For example, $B_0(s)$ has the slopes at s_0 and s_1 and we are going to add the phase of the crossovers at s_2, s_3 and s_4.

First we define an allpass filter $A_n(s)$ that exhibits a flat magnitude response and the phase shift of a $(2n)^{\text{th}}$ order Butterworth lowpass as

$$A_n(s_c) = \frac{D_n^*(s_c)}{D_n(s_c)}\,, \tag{5.13}$$

where D_n is the n^{th} order Butterworth polynomial and $(.)^*$ denotes complex conjugation.

The allpass filterbank $\tilde{A}_j(s)$ is then given by

$$\tilde{A}_0(s) = A_3(s_2)\, A_3(s_3)\, A_4(s_4)\,, \tag{5.14}$$
$$\tilde{A}_1(s) = A_2(s_0)\, A_3(s_3)\, A_4(s_4)\,, \tag{5.15}$$
$$\tilde{A}_2(s) = A_2(s_0)\, A_2(s_1)\, A_4(s_4)\,, \tag{5.16}$$
$$\tilde{A}_3(s) = A_2(s_0)\, A_3(s_1)\, A_3(s_2)\,, \tag{5.17}$$
$$\tilde{A}_4(s) = A_2(s_0)\, A_3(s_1)\, A_3(s_2)\, A_3(s_3) \tag{5.18}$$

and the phase corrected bandpass filterbank $\tilde{B}_j(s)$, that sums up to a nearly flat magnitude response (see. Fig. 5.4), denotes to

$$\tilde{B}_j(s) = B_j(s)\, \tilde{A}_j(s) \quad j = 0, \ldots, 4\,. \tag{5.19}$$

The sum of these bandpass filters might deviate in magnitude from a perfectly flat response, as only single crossover pairs are complementary. Deviations are corrected however by a minimum-phase filter (see Sec. 5.6). We discretize the frequency response and introduce the discrete frequency bin $k = 1, ..., \mathrm{N_{FFT}}$.

To ensure constant amplitude of radiated sound (acoustic free-field normalization) across the bands of the filterbank we need to normalize the weights

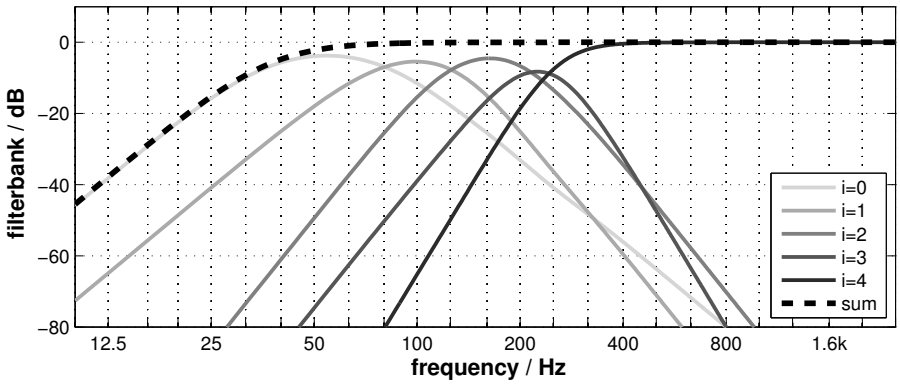


Figure 5.4.: Filterbank for the regularization of the radial filters.

applied to the spherical harmonics. We compute the normalization factors c_j, that depend on the max-$\boldsymbol{r}_\mathrm{E}$ weights $w_{n,\mathrm{N}}$ as follows

$$w_{n,\mathrm{N}} = P_n\big[\cos\big(\tfrac{\pi}{180}\tfrac{137.9}{\mathrm{N}+1.51}\big)\big] \quad n = 0,\ldots,\mathrm{N} \tag{5.20}$$

$$c_j = \sum_{n=0}^{j} w_{n,j}\,(2n+1) \quad j = 0,\ldots,4 \tag{5.21}$$

The band-dependent weights are stacked into the matrix $[w_{n,j}]$ to define the weight vectors $\boldsymbol{w}_j$:

$$[w_{n,j}] = \begin{bmatrix} w_{0,0} & w_{0,1} & w_{0,2} & w_{0,3} & w_{0,4} \\ 0 & w_{1,1} & w_{1,2} & w_{1,3} & w_{1,4} \\ 0 & 0 & w_{2,2} & w_{2,3} & w_{2,4} \\ 0 & 0 & 0 & w_{3,3} & w_{3,4} \\ 0 & 0 & 0 & 0 & w_{4,4} \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_0 & \boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3 & \boldsymbol{w}_4 \end{bmatrix}. \tag{5.22}$$

Applying the normalization yields the weight vectors $\tilde{\boldsymbol{w}}_j$ for a constant amplitude across all bands

$$\tilde{\boldsymbol{w}}_j = \frac{\boldsymbol{w}_j}{c_j} \quad j = 0,\ldots,4\,. \tag{5.23}$$

We define the vector of non-regularized far-field radial filters H_n^{-1}, which are depicted in Fig. 5.3,

$$\boldsymbol{h}_{\mathrm{inv}}(k) = [H_n(k)^{-1}]_{n=0\ldots4} \quad H_n(k) = \frac{\rho_0 c}{\tilde{k}} \frac{i^n}{h_n'(\tilde{k}\mathrm{R}) e^{i\tilde{k}\mathrm{R}}} \,, \tag{5.24}$$

with the wave number $\tilde{k}$ relating to the discrete frequency bin $k = 1, ..., \mathrm{N_{FFT}}$

$$\tilde{k} = \frac{2\pi}{c} \frac{k}{\mathrm{N_{FFT}}} f_s \,, \tag{5.25}$$

and a vector that consists of the order-dependent cap model coefficients a_n

$$\boldsymbol{a} = [a_0, \ldots, a_4] \,, \tag{5.26}$$

to compute the j^{th} band of regularized radial filters:

$$\boldsymbol{f}_j(k) = \tilde{B}_j(k)\, \boldsymbol{h}_{\mathrm{inv}}(k)\, \mathrm{diag}\{\tilde{\boldsymbol{w}}_j\} \mathrm{diag}\{\boldsymbol{a}\}^{-1} \,. \tag{5.27}$$

Lastly, we sum the bands together (as it is a parallel section of Linkwitz-Riley crossovers) and end up with order dependent regularized radial filters as depicted in Fig. 5.5.

$$\boldsymbol{f}(k) = \sum_{j=0}^{4} \boldsymbol{f}_j(k) = [f_n(k)]_{n=0,\ldots,4} \,. \tag{5.28}$$

We expand it to account for the $(2n+1)$ degrees within an order n and apply our mixed-order projection matrix $\boldsymbol{M}$:

$$\boldsymbol{f}_{\mathrm{N}}(k) = [f_n(k)]_{n=0\ldots\mathrm{N},m=-n\ldots n} \,, \tag{5.29}$$

$$\boldsymbol{f}_{\mathrm{M}}(k) = \boldsymbol{M}\, \boldsymbol{f}_{\mathrm{N}}(k) \,. \tag{5.30}$$

To verify the radial filters and the cut-off frequencies we plot the maximum excursion in dB for every band and for the sum of all bands, see Fig. 5.6. Therefore, we encode $B = 120$ beam directions (t-design), apply the radial

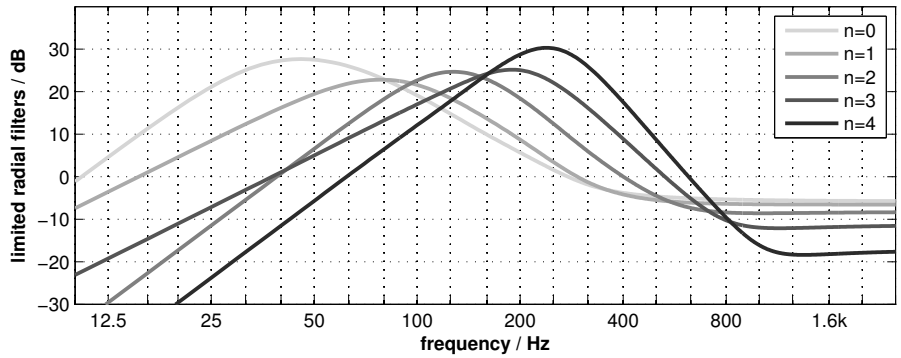


Figure 5.5.: Limited radial filters.

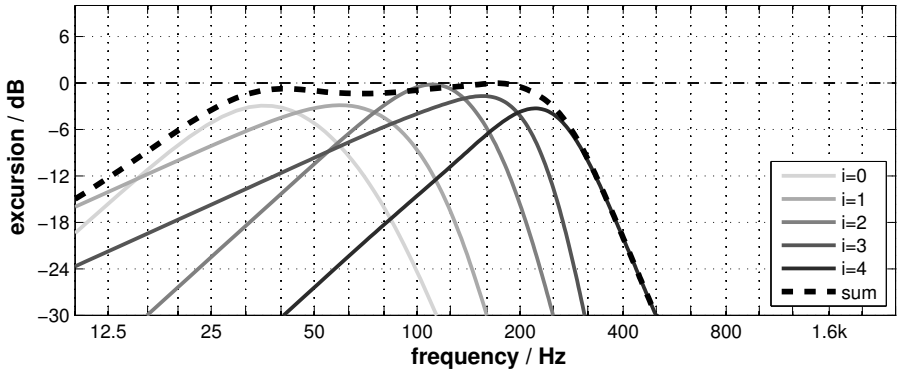


Figure 5.6.: Limited driver excursion.

filterbank and decode to L velocities. By multiplying with $1/(i2\pi\frac{k}{\mathrm{N_{FFT}}}f_s)$ we integrate the velocity $v_{\mathrm{max,j}}(k)$ and get the excursion $x_{\mathrm{max,j}}(k)$

$$v_{\mathrm{max},j}(k) = \max\{\boldsymbol{Y}_{\mathrm{M}}^{+}\,\mathrm{diag}\{\boldsymbol{f}_j(k)\}\,\boldsymbol{Y}_{beam}\}\,, \tag{5.31}$$

$$x_{\mathrm{max},j}(k) = \frac{v_{\mathrm{max},j}(k)}{i2\pi\frac{k}{\mathrm{N_{FFT}}}f_s}\,, \tag{5.32}$$

$$v_{\mathrm{max}}(k) = \max\{\boldsymbol{Y}_{\mathrm{M}}^{+}\,\mathrm{diag}\{\boldsymbol{f}(k)\}\,\boldsymbol{Y}_{beam}\}\,, \tag{5.33}$$

$$x_{\mathrm{max}}(k) = \frac{v_{\mathrm{max}}(k)}{i2\pi\frac{k}{\mathrm{N_{FFT}}}f_s}\,. \tag{5.34}$$

5.4. MIMO Crosstalk-Cancellation

By mounting the loudspeakers in a common enclosure, the large air volume represents less stiffness at low frequencies, but we face acoustic crosstalk. That is, if one transducer is driven by a signal, the others will vibrate passively. This is undesirable, as we need to control the transducers independently to achieve modal beamforming. Formally, our L × L MIMO system $\boldsymbol{T}$ can be described as

$$\boldsymbol{v}(\omega) = \boldsymbol{T}(\omega)\, \boldsymbol{u}(\omega). \tag{5.35}$$

In the following, the frequency dependence of all MIMO and SISO systems is dropped to increase readabiltiy.

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_\mathrm{L} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1\mathrm{L}} \\ T_{21} & T_{22} & \dots & T_{2\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots \\ T_{\mathrm{L}1} & T_{\mathrm{L}2} & \dots & T_{\mathrm{LL}} \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_\mathrm{L} \end{bmatrix} \tag{5.36}$$

A system inversion yields the voltage signals $\boldsymbol{u}$ for a decoupled control of cone velocities.

$$\boldsymbol{u} = \boldsymbol{T}^{-1} \cdot \boldsymbol{v} \tag{5.37}$$

A full system inversion, that results in both flat magnitude responses of the direct paths and crosstalk-cancellation over the whole frequency range can lead to acausal filters and unfeasibly long impulse responses in the time domain. Regarding the crosstalk-cancellation, we can effectively reduce the complexity of the inverted system by discarding frequencies above and below certain cut-off frequencies. For the direct paths, we only equalize the drivers to a mean response. A flat response can be created in the very end of the control chain and is not necessary to be achieved by the MIMO canceller.

The active responses equalized to the mean response H_mean yield the SISO equalizers $H_{\mathrm{eq},l}$,

$$\begin{aligned} T_{ll} H_{\mathrm{eq},l} &= H_\mathrm{mean}, \qquad \text{for } l = 1, \dots, \mathrm{L}, \\ \boldsymbol{h}_\mathrm{eq} &= [H_{\mathrm{eq},l}]_{l=1,\dots,\mathrm{L}} \end{aligned} \tag{5.38}$$

that are applied to get the equalized MIMO system

$$\begin{aligned}\boldsymbol{T}_{\text{eqd}} &= \boldsymbol{T}\ \text{diag}\{\boldsymbol{h}_{\text{eq}}\} \\ &= \begin{bmatrix} H_{\text{mean}} & T_{12}H_{eq,2} & \dots & T_{1\text{L}}H_{eq,\text{L}} \\ T_{21}H_{eq,1} & H_{\text{mean}} & \dots & T_{2\text{L}}H_{eq,\text{L}} \\ \vdots & \vdots & \ddots & \vdots \\ T_{\text{L}1}H_{eq,1} & T_{\text{L}2}H_{eq,2} & \dots & H_{\text{mean}} \end{bmatrix} \\ &= H_{\text{mean}}\boldsymbol{I} + [\boldsymbol{T}_{\text{eqd}} - \text{diag}\{\text{diag}\{\boldsymbol{T}_{\text{eqd}}\}\}] \\ &= \boldsymbol{T}_{\text{eqd,active}} + \boldsymbol{T}_{\text{eqd,passive}}\ .\end{aligned} \tag{5.39}$$

To obtain short crosstalk cancellation filters, regularization is applied. As mentioned, the filter complexity can be reduced significantly by bandpass filtering the passive system matrix before the inversion.

The lower and upper cut-off frequencies of the bandpass determine the range that we wish to achieve crosstalk-cancellation.

Passive paths are linear-phase band-passed for regularization

$$\tilde{\boldsymbol{T}}_{\text{eqd}} = \boldsymbol{T}_{\text{eqd,active}} + \boldsymbol{T}_{\text{eqd,passive}}\ H_{\text{BP}}\ , \tag{5.40}$$

and altogether, the SISO equalizers times the filtered inverse $\tilde{\boldsymbol{T}}_{\text{eqd}}^{-1}$ times the mean response yields a matching and crosstalk cancelling system $\boldsymbol{X}_c$

$$\boldsymbol{X}_c = \text{diag}\{\boldsymbol{h}_{\text{eq}}\}\ \tilde{\boldsymbol{T}}_{\text{eqd}}^{-1}\ H_{\text{mean}}. \tag{5.41}$$

Ideally, in the frequency range of the bandpass ($H_{\text{BP}} = 1$, so $\tilde{\boldsymbol{T}}_{\text{eqd}} = \boldsymbol{T}_{\text{eqd}}$) this equalization yields a crosstalk-cancelled system matching the active responses, applied from the voltage side:

$$\begin{aligned}\boldsymbol{T}\boldsymbol{X}_c &= \boldsymbol{T}\ \text{diag}\{\boldsymbol{h}_{\text{eq}}\}\ \tilde{\boldsymbol{T}}_{\text{eqd}}^{-1}\ H_{\text{mean}} \\ &= \boldsymbol{T}\ \text{diag}\{\boldsymbol{h}_{\text{eq}}\}\text{diag}\{\boldsymbol{h}_{\text{eq}}\}^{-1}\ \boldsymbol{T}^{-1}\ H_{\text{mean}} \\ &= \begin{bmatrix} H_{mean} & 0 & \dots & 0 \\ 0 & H_{mean} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & H_{mean} \end{bmatrix}.\end{aligned} \tag{5.42}$$

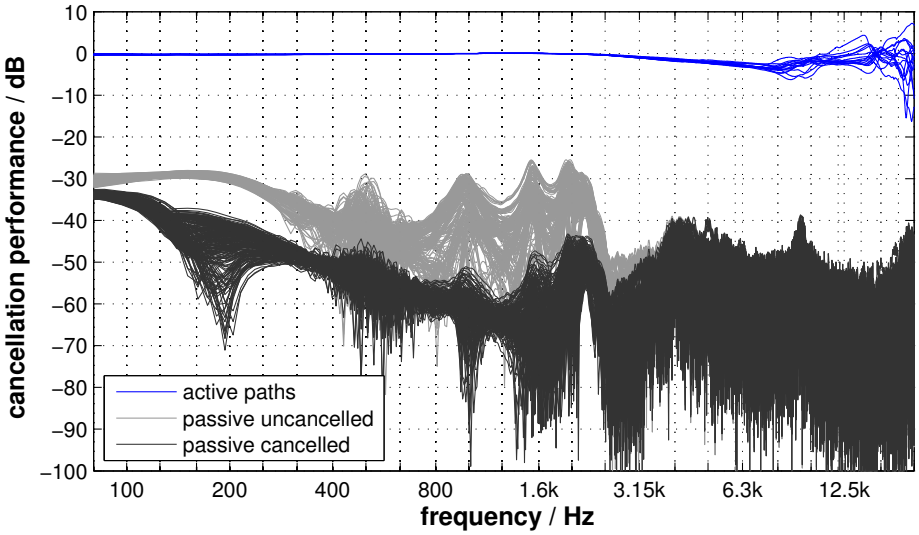


Figure 5.7.: Crosstalk-Cancellation performance for the 3|9|3-array, all curves relative to mean active response. Cut-off frequencies $f_{b1} = 100$ Hz and $f_{b2} = 2.9$ kHz. Active paths are shown in blue. Passive paths in grey.

Fig. 5.7 shows the crosstalk-cancellation performance for the 3|9|3-array. The mean active path has been subtracted to show that within the bandpass frequency range ($f_{\mathrm{b1}} = 100$ Hz to $f_{\mathrm{b2}} = 2.9$ kHz) we have little deviation between active responses and significant cancellation of up to 20 dB. The bandpass regularization of the crosstalk paths before system inversion has a beneficial effect on impulse response length of the cancellation filters as shown in Fig. 5.8.

5.5. High-Frequency AllRAD Panning

In the high-frequency band we employ AllRAD-based panning that approximates VBAP while remaining in the ambisonics domain (see Subsec. 3.3.2). The decoder was derived in Eq. (3.30) to

$$\boldsymbol{D} = \frac{4\pi}{\mathrm{J}}\, \boldsymbol{G}\, \boldsymbol{Y}_{\mathrm{N,J}}\, \mathrm{diag}\{\boldsymbol{w}_{\mathrm{N}}\}\ . \tag{5.43}$$

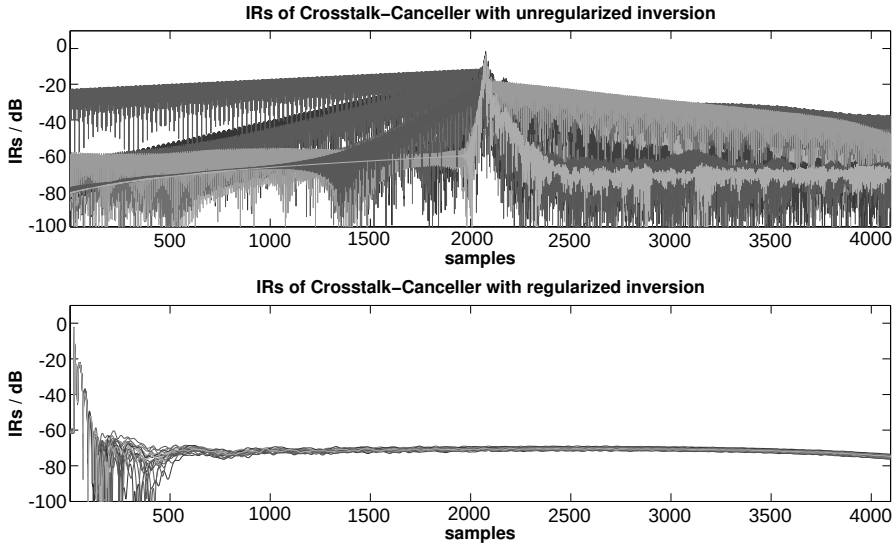


Figure 5.8.: Bandpass filtering of passive paths before the inversion (regularized inversion) yields short and causal direct-path impulse responses for the crosstalk-canceller.

To compensate the phase delay of the lower band, that is caused by the crosstalk canceller and the Linkwitz-Riley crossover a simple group delay has to be applied to the high-frequency band. The amount of samples n_d is calculated by cross-correlation and yields $d(k) = \exp\{-\mathrm{i}k\frac{2\pi}{N_{\mathrm{FFT}}}n_d\}$.

5.6. Band Summation and Timbral Equalization

For every frequency bin k we evaluate the following matrix computations

$$\boldsymbol{H}_1(k) = H_{\mathrm{HP}}(k)\ \boldsymbol{D}\ d(k)\,, \tag{5.44}$$

$$\boldsymbol{H}_2(k) = H_{\mathrm{LP}}(k)\ \boldsymbol{X}_c(k)\ \boldsymbol{Y}_{\mathrm{M}}^{+}\ \mathrm{diag}\{\boldsymbol{f}(k)\}\,, \tag{5.45}$$

$$\boldsymbol{H}(k) = [\boldsymbol{H}_1(k) + \boldsymbol{H}_2(k)]\ E(k)\ . \tag{5.46}$$

The last block in the processing chain is the timbral on-axis beam equalization $E(k)$. It flattens the response of the mean of all on-axis beams (beams directed

towards one of the loudspeakers). This equalization does not influence the beam pattern, but linearizes the timbral quality of the loudspeaker array.

It is based on directional impulse response (DIR) measurements from all of the L loudspeakers to a surrounding spherical microphone array with an effective number of $R = 648$ measurement points, where each path of the MIMO system is described by an impulse response of $n_{\mathrm{IR}} = 1024$ samples.

The idea is to combine the measurements with our specific control filters $\boldsymbol{H}$ and thus have access to the magnitude response of the radiated beams. A filter is then designed that equalizes the response.

The definition of the following matrices is helpful:

$$\boldsymbol{P}(k) = [P_{rl}(k)]_{r=1\ldots R}^{l=1\ldots \mathrm{L}} \qquad [R \times \mathrm{L}] \tag{5.47}$$

$$\boldsymbol{Y}_{\mathrm{mic}} = [Y_n^m(\boldsymbol{\theta}_r)]_{r=1\ldots R}^{n=0\ldots \mathrm{N_{mic}}, m=-n\ldots n} \qquad [R \times (\mathrm{N_{mic}}+1)^2] \tag{5.48}$$

$$\boldsymbol{Y}_{\mathrm{L}} = [Y_n^m(\boldsymbol{\theta}_l)]_{n=0\ldots \mathrm{N}, m=-n\ldots n}^{l=1\ldots \mathrm{L}} \qquad [(\mathrm{N}+1)^2 \times \mathrm{L}] \tag{5.49}$$

$$\boldsymbol{Y}_{\mathrm{eval}} = [Y_n^m(\boldsymbol{\theta}_l)]_{l=1\ldots \mathrm{L}}^{n=0\ldots \mathrm{N_{mic}}, m=-n\ldots n} \qquad [\mathrm{L} \times (\mathrm{N_{mic}}+1)^2]\,, \tag{5.50}$$

where $\boldsymbol{P}(k)$ is the $R \times \mathrm{L}$ MIMO measurement matrix in the frequency domain at the frequency bin k and $\boldsymbol{Y}_{\mathrm{mic}}$ is a matrix of spherical harmonics sampled at the R microphone positions up to the analysis order $\mathrm{N_{mic}} = 17$. $\boldsymbol{Y}_{\mathrm{L}}$ and $\boldsymbol{Y}_{\mathrm{eval}}$ are the on-axis beam-encoding and on-axis sound-pressure evaluating matrices, sampled up to the orders N and $\mathrm{N_{mic}}$ respectively.

First, the DIR measurements are encoded into the spherical harmonics domain with an analysis order $\mathrm{N_{mic}} = 17$

$$\boldsymbol{P}_{\mathrm{SH}}(k) = \boldsymbol{Y}_{\mathrm{mic}}^{+}\, \boldsymbol{P}(k)\,. \tag{5.51}$$

Now we combine the control filters $\boldsymbol{H}(k)$ with the measurements by frequency-domain MIMO filtering to obtain a $(\mathrm{N_{mic}}+1)^2 \times (\mathrm{N}+1^2)$ system $\boldsymbol{W}$ that describes the transmission from an input pattern of $(\mathrm{N}+1)^2$ spherical harmonics to an acoustically measured pressure spectrum of $(\mathrm{N_{mic}}+1)^2$ spherical harmonics

$$\boldsymbol{W}(k) = \boldsymbol{P}_{\mathrm{SH}}(k)\, \boldsymbol{H}(k)\ . \tag{5.52}$$

We can apply far-field extrapolation filters to the system $\boldsymbol{W}$ (measured pressure spectrum at $\mathrm{R} = 0.75$ m)

$$\boldsymbol{h}(k) = \left[\frac{i^{n+1}}{\tilde{k}\, h_n(\tilde{k}\mathrm{R})\, e^{i\tilde{k}\mathrm{R}}}\right]_{n=0\ldots\mathrm{N_{mic}}, m=-n\ldots n} \qquad \tilde{k} = \frac{2\pi}{c}\frac{k}{\mathrm{N_{FFT}}} f_s\,, \tag{5.53}$$

$$\boldsymbol{W}_{\mathrm{r}\gg}(k) = \mathrm{diag}\{\boldsymbol{h}(k)\}\, \boldsymbol{W}(k)\ . \tag{5.54}$$

Because we have a fully characterized system in the SH-domain, we can investigate any combination of encoded beam directions and analysis directions. As we are interested in the frequency response of the L on-axis beams we encode these directions into a $(\mathrm{N}+1)^2 \times \mathrm{L}$ matrix $\boldsymbol{Y}_{\mathrm{L}}$ and right-multiply to our transmission system $\boldsymbol{W}_{\mathrm{r}\gg}$.

Finally, we evaluate the pressure spectrum at the directions of the loudspeakers to obtain a matrix of $\mathrm{L} \times \mathrm{L}$ frequency responses (L on-axis beams to the identical L directions of interest)

$$\boldsymbol{P}_{\mathrm{L}\times\mathrm{L},\mathrm{r}\gg}(k) = \boldsymbol{Y}_{\mathrm{eval}}\, \boldsymbol{W}_{\mathrm{r}\gg}(k)\, \boldsymbol{Y}_{\mathrm{L}}\,. \tag{5.55}$$

With this $\mathrm{L} \times \mathrm{L}$ transmission matrix we can compute either a diffuse-field response or a free-field response (using only the diagonal elements of the MIMO matrix). The free-field response $P_{\mathrm{FF}}(k)$ is calculated by averaging over all the on-axis responses

$$P_{\mathrm{FF}}(k) = \sqrt{\frac{1}{\mathrm{L}}\sum_{l=1}^{\mathrm{L}} |P_{ll}(k)|^2}\,. \tag{5.56}$$

For the diffuse-field response we average twice, first for every beam direction over the L output paths. After that, we average the L diffuse beam responses to a single diffuse-field response $P_{\mathrm{DF}}(k)$

$$P_{\mathrm{DF}}(k) = \sqrt{\frac{1}{\mathrm{L}^2} \sum_{l=1}^{\mathrm{L}} \sum_{j=1}^{\mathrm{L}} |P_{lj}(k)|^2} \,. \tag{5.57}$$

As our equalizer $E(k)$ we choose the regularized inverse of the 12^{th}-octave smoothed free-field response $P_{\mathrm{FF}}(k)$, as depicted in Fig. 5.9. The frequency response is reconstructed as a minimum-phase response from the desired magnitude response. We therefore compute the real cepstrum $\tilde{p}$ which is an even real-valued signal, as we apply an inverse DFT to a real magnitude spectrum

$$\tilde{p}[n] = \mathcal{IDFT}\{\ln|E(k)|\} \,. \tag{5.58}$$

Thus, to make the frequency response minimum-phase, we create a purely causal cepstrum by applying a window function that effectively adds the acausal cepstral coefficients to the causal ones (flip non-minimum phase zeros into the unit circle) without altering the magnitude response. [15]

$$\tilde{p}_{\mathrm{causal}}[n] = \begin{cases} 2 \cdot \tilde{p}[n] & 1 \leq n \leq \mathrm{N_{FFT}}/2 - 1 \\ \tilde{p}[n] & n = 0 \ \wedge \ n = \mathrm{N_{FFT}}/2 \\ 0 & n < 0 \end{cases} , \tag{5.59}$$

$$\text{for} \quad n = [0, \ldots, \mathrm{N_{FFT}}/2, -\mathrm{N_{FFT}}/2 + 1, ..., -1] \,. \tag{5.60}$$

We obtain the minimum-phase impulse response by the following inverse transformations:

$$E_{\mathrm{min.ph.}}[k] = \exp\Big(\mathcal{DFT}\{\,\tilde{p}_{\mathrm{causal}}[n]\,\}\Big) \,, \tag{5.61}$$

$$e_{\mathrm{min.ph.}}[n] = \mathcal{IDFT}\{\,E_{\mathrm{min.ph.}}[k]\,\} \,. \tag{5.62}$$

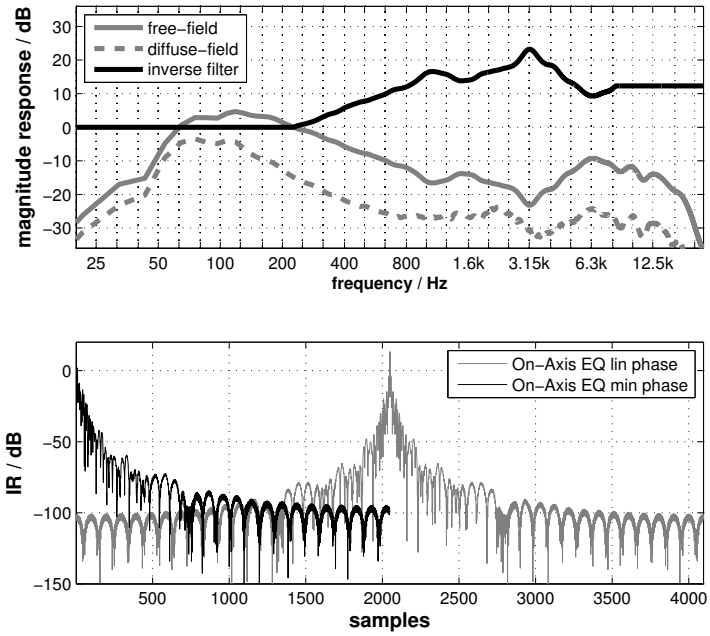


Figure 5.9.: On-axis timbral beam equalization. A regularized inverse magnitude response (black) is created from the free-field response (grey). By making the inverse filter minimum-phase we get a low-latency impulse response (black) in the time-domain.

6. Directivity Plots and Array Comparison

In this chapter a measurement-based evaluation of beam patterns is conducted. Recalling the definition of the system $\boldsymbol{W}(k)$ from Sec. 5.6,

$$\boldsymbol{P}_{\mathrm{SH}}(k) = \boldsymbol{Y}_{\mathrm{mic}}^{+}\, \boldsymbol{P}(k) \tag{6.1}$$
$$\boldsymbol{W}(k) = \boldsymbol{P}_{\mathrm{SH}}(k)\, \boldsymbol{H}(k)\,, \tag{6.2}$$

we can investigate any combination of encoded beam direction and analysis directions. As we are interested mostly in horizontal beams, we encode a beam of horizontal direction $\boldsymbol{\theta}_b$ and evaluate it by expanding the pressure spectrum on a horizontal or vertical cut with a resolution of one degree, resulting in a 360×1 sound pressure vector $\boldsymbol{p}(k)$

$$\boldsymbol{p}(k) = \boldsymbol{Y}_{\mathrm{eval}}\, \boldsymbol{W}(k)\, \boldsymbol{y}_{\mathrm{b}}\,, \tag{6.3}$$

using the following matrices

$$\boldsymbol{P}(k) = [P_{rl}(k)]_{r=1\ldots R}^{l=1\ldots L} \qquad [R \times L] \tag{6.4}$$
$$\boldsymbol{Y}_{\mathrm{mic}} = [Y_n^m(\boldsymbol{\theta}_r)]_{r=1\ldots R}^{n=0\ldots \mathrm{N}_{\mathrm{mic}}, m=-n\ldots n} \qquad [R \times (\mathrm{N}_{\mathrm{mic}}+1)^2] \tag{6.5}$$
$$\boldsymbol{y}_{\mathrm{b}} = [Y_n^m(\boldsymbol{\theta}_b)]_{n=0\ldots \mathrm{N}, m=-n\ldots n} \qquad [(\mathrm{N}+1)^2 \times 1] \tag{6.6}$$
$$\boldsymbol{Y}_{\mathrm{eval}} = [Y_n^m(\boldsymbol{\theta}_j)]_{j=1\ldots 360}^{n=0\ldots \mathrm{N}_{\mathrm{mic}}, m=-n\ldots n} \qquad [360 \times (\mathrm{N}_{\mathrm{mic}}+1)^2]\,, \tag{6.7}$$
$$\tag{6.8}$$

where $\boldsymbol{Y}_{\text{eval}}$ is now the matrix of spherical harmonics used to evaluate horizontal/vertical cuts. As we are interested in the directivity of a beam we compute the squared magnitude of each element of the complex-valued pressure vector $\boldsymbol{p}(k)$ and normalize for every frequency bin k to display a directivity in relative decibels as seen in the following plots.

$$p_{\text{plot},j}(k) = 20 \cdot \log_{10}\left(\frac{|p_j(k)|}{|p_{\max}(k)|}\right) \quad j = 1, ..., 360 \tag{6.9}$$

The following plots compare the new mixed-order plus AllRAD control with traditional control, based on the acoustic MIMO measurements, see Sec. 4.4. First we investigate two platonic arrays, a low-cost dodecahedral array, seen in Fig. 6.1, and the IEM IKO2 icosahedral array in Fig. 6.2. It is visible that beams become significantly narrower in the range of 300 to 800 Hz and modal break-up in the high-frequency range is less prominent. For the new mixed-order arrays the AllRAD approach shows improvements as horizontal beam directions can be on-axis loudspeaker directions (Fig. 6.3). A comparison of on-axis panning versus in-between two loudspeakers is shown in Fig. 6.7 for the 393-array and reveals expectable side-lobes that effectively widen the beam pattern above 4 kHz.

Comparing the different arrays, we can see that the smaller 3|9|3-array is octave-shifted in its directivity, namely achieves 4^{th} order beamforming between 500 and 2000 Hz and that the new ring layouts generally prevent a break-up of the beam pattern for high-frequencies (see horizontal and vertical cuts in Figs. 6.5 and 6.6). Informal listening tests confirmed this to have an audible, beneficial effect on the localization of broadband signals.

Finally, with Fig. 6.8 it is proven that the erroneous vertical off-axis radiation at high frequencies and horizontal beam directions is not caused by the AllRAD approach, but is inherent to platonic layouts.

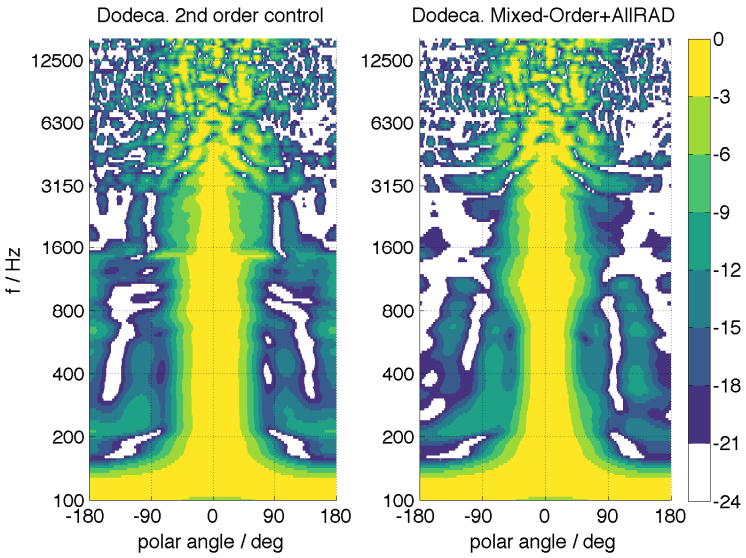


Figure 6.1.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Crossover to AllRAD panning at 1 kHz.

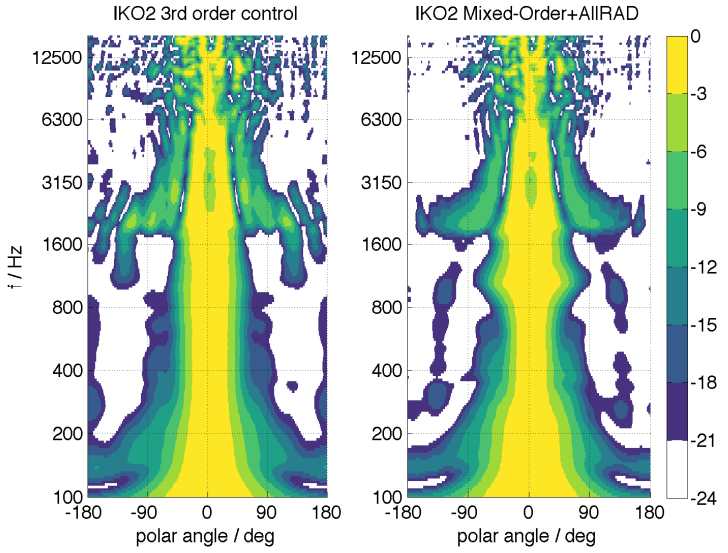


Figure 6.2.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Crossover to AllRAD panning at 1 kHz.

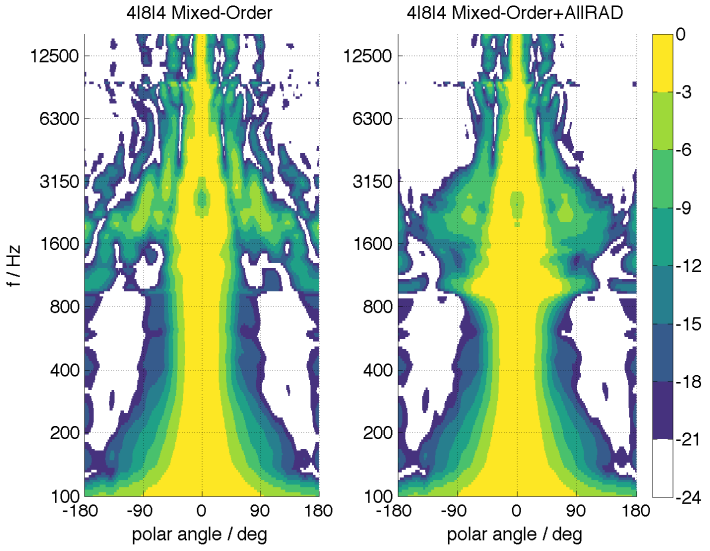


Figure 6.3.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Crossover to AllRAD panning at 1 kHz.

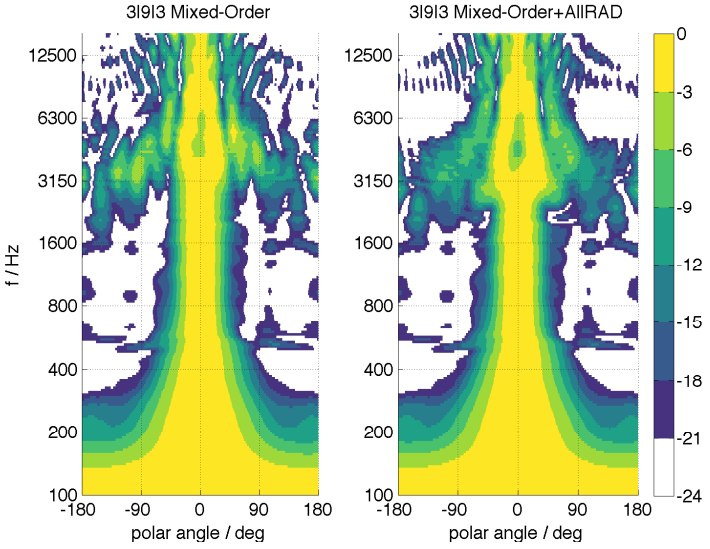


Figure 6.4.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Crossover to AllRAD panning at 2.9 kHz.

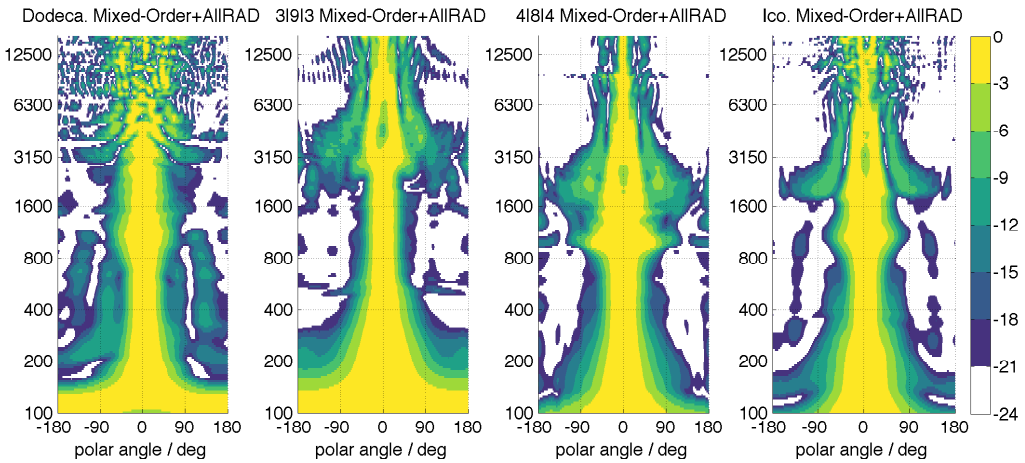


Figure 6.5.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Sorted by increasing number of transducers from left to right. The 4|8|4-layout significantly reduces side-lobes at high frequencies, while saving 4 loudspeakers compared to an icosahedral array.

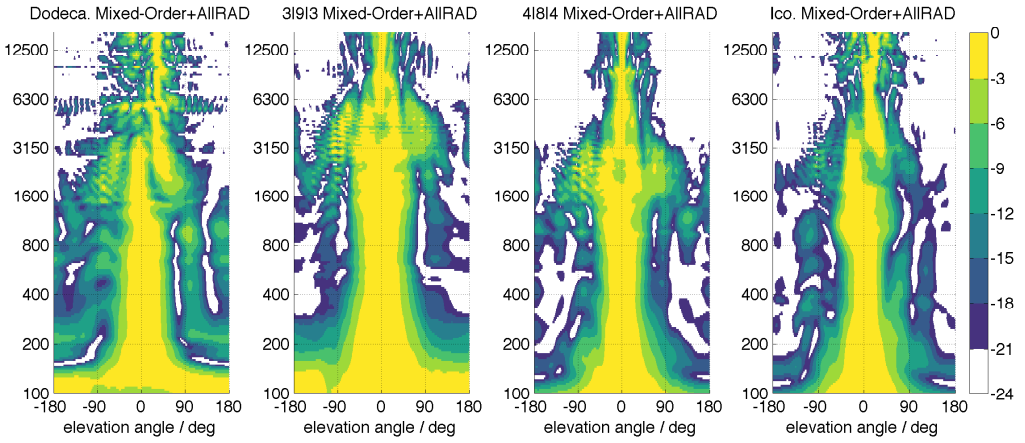


Figure 6.6.: Vertical cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency. Sorted by increasing number of transducers from left to right. Mixed-order ring layouts prevent the (vertical) off-axis radiation for horizontal beams at high frequencies.

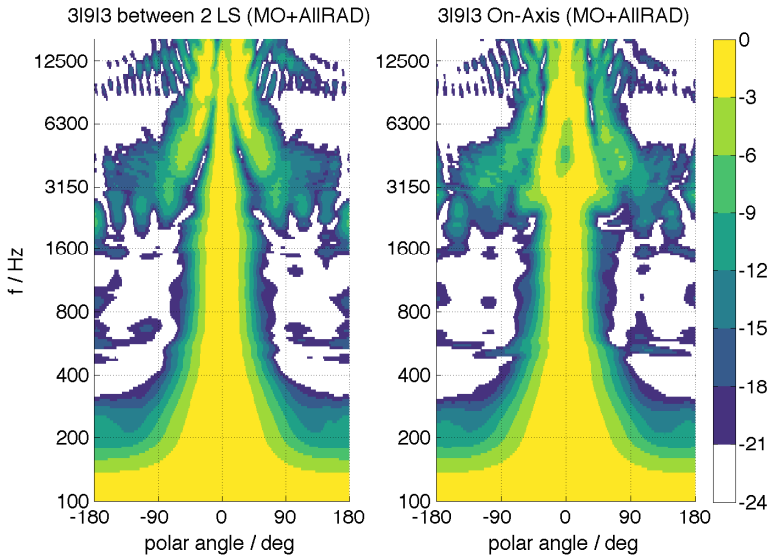


Figure 6.7.: Horizontal cut of horizontal beams $\varphi = 20°$ vs. $\varphi = 0°$, $\theta = 90°$. Panning directions in between two loudspeakers (left) cause side-lobes above 4 kHz in comparison to on-axis panning directions (right).

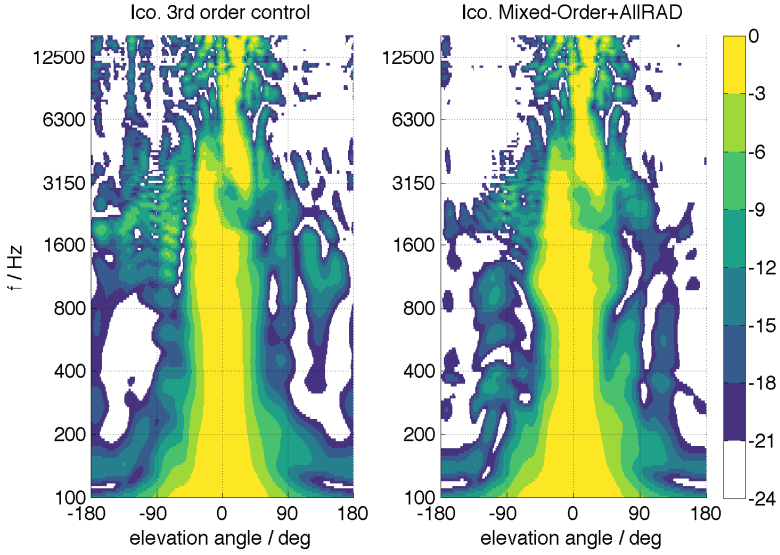


Figure 6.8.: Vertical cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. The off-axis radiation for horizontal beams at high frequencies is not due to AllRAD panning, but inherent to platonic layouts.

7. Conclusion

This thesis presented mixed-order control and high-frequency panning that are meant to enrich beamforming with compact spherical loudspeaker arrays. To evaluate the improvements, effective horizontal 2D and global 3D order measures were introduced that confirmed that the horizontal directivity of the dodecahedron can be increased from an effective second to third order, and for the icosahedron from third to fourth, with only negligible impact on the effective 3D directivity order.

New mixed-order layouts were introduced that are composed of three loudspeaker rings. The dedicated mixed-order layouts save transducers while achieving equal or higher beam orders in the horizontal plane. They are especially suited for the AllRAD panning as horizontal panning directions can be on-axis loudspeaker directions.

This work is a first step towards manufacturing affordable electroacoustic instruments using accurate spherical beamforming for both home-studio rehearsal and electroacoustic chamber music performances. Material cost and manufacturing time can be significantly reduced through 3D-printed housings and a reduced number of smaller, but high quality loudspeakers with a frequency range from 100 Hz to 20 kHz. An overall material cost below 800 euros and an overall weight of three kilograms make the 3|9|3-array prototype a highly mobile and affordable beamformer instrument.

Further work should consider designing a suitable and compact amping/interface solution that enables easy transportation of the overall system. Lastly, a psychoacoustic evaluation of the new arrays would be interesting regarding the overall increased directivity for horizontal beam directions, which enabled some (typically broadband) sounds to be localized even in room corners behind the listener during first informal listening experiments.

Bibliography

[1] Rimas Avizienis et al. “A compact 120 independent element spherical loudspeaker array with programable radiation patterns.” In: *Audio Eng. Soc. Convention 120*. 2006.

[2] C Bradford Barber, David P Dobkin, and Hannu Huhdanpaa. “The quickhull algorithm for convex hulls.” In: *ACM Transactions on Mathematical Software (TOMS)* 22.4 (1996), pp. 469–483.

[3] Jiho Chang and Marton Marschall. “Periphony-Lattice Mixed-Order Ambisonic Scheme for Spherical Microphone Arrays.” In: *IEEE/ACM TASLP* 26.5 (2018), pp. 924–936.

[4] Jérôme Daniel. “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia.” In: (2001).

[5] Thomas Deppisch et al. “Surround with Depth on First-Order Beam-Controlling Loudspeakers.” In: *Audio Eng.Soc. Convention 144*. Audio Engineering Society. 2018.

[6] M. Graef. URL: `https://homepage.univie.ac.at/manuel.graef/quadrature.php`.

[7] Stefan Lösler. “MIMO-Rekursivfilter für Kugelarrays.” In: *Master’s thesis, Univ. of Music and Perf. Arts, Graz, Austria* (2014).

[8] Márton Marschall and Jörg Buchholz. “Capturing and reproducing realistic acoustic scenes for hearing research.” PhD thesis. TU Denmark, Dep. of Elec.Eng., 2014.

[9] Jens Meyer and Gary Elko. “Spherical harmonic modal beamforming for an augmented circular microphone array.” In: *IEEE ICASSP*. 2008, pp. 5280–5283.

[10] Nils Meyer-Kahlen, Franz Zotter, and Katharina Pollack. "Design and Measurement of a First-Order, Horizontally Beam-Controlling Loudspeaker Cube." In: *Audio Engineering Society Convention 144*. Audio Engineering Society. 2018.

[11] Alexander Mattioli Pasqual et al. "Application of acoustic radiation modes in the directivity control by a spherical loudspeaker array." In: *Acta acustica united with Acustica* 96.1 (2010), pp. 32–42.

[12] Martin Pollow and Gottfried K Behler. "Variable directivity for platonic sound sources based on spherical harmonics optimization." In: *Acta Acustica united with Acustica* 95.6 (2009), pp. 1082–1092.

[13] H. Pomberger. "Angular and Radial Directivity Control for Spherical Loudspeaker Arrays." In: *Diploma Thesis*. Univ. of Music and Perf. Arts Graz, 2008.

[14] Ville Pulkki. "Virtual sound source positioning using vector base amplitude panning." In: *Journal of the Audio Eng. Soc.* 45.6 (1997), pp. 456–466.

[15] Julius O. Smith. *Introduction to Digital Filters with Audio Applications*. online book. `http://ccrma.stanford.edu/~jos/filters/`, accessed on 01/15/19.

[16] Olivier Warusfel, Philippe Derogis, and Rene Causse. "Radiation synthesis with digitally controlled loudspeakers." In: *Audio Eng. Soc. Convention 103*. 1997.

[17] Robert Webb. *Robert Webb's Stella software*. URL: `http://www.software3d.com/Stella.php`.

[18] Earl G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Acad.Press, 1999.

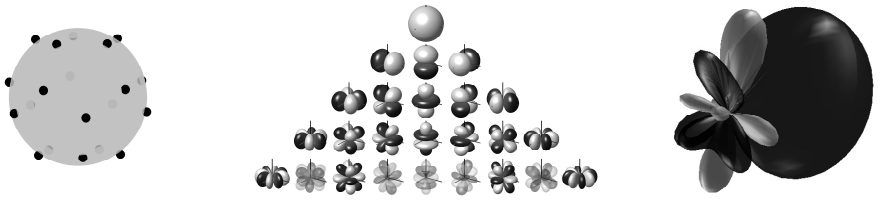
[19] F. Zotter and M. Frank. "All-Round Ambisonic Panning and Decoding." In: *Journal of the Audio Eng. Soc.* 60.10 (2012), pp. 801–820.

[20] F. Zotter and R. Hoeldrich. "Modeling Radiation Synthesis with Spherical Loudspeaker Arrays." In: *Proc. of the 19th Int. Congress on Acoustics*. Madrid, 2007.

[21] Franz Zotter et al. “A beamformer to play with wall reflections: The icosahedral loudspeaker.” In: *Computer Music Journal* 41.3 (2017), pp. 50–68.

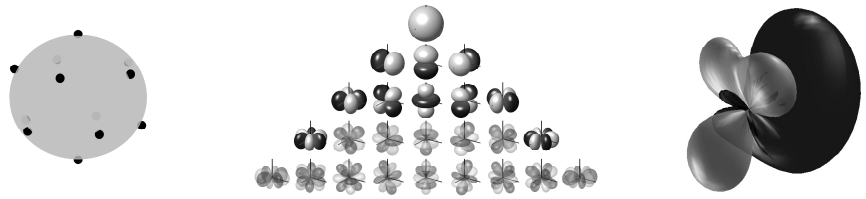
Appendix A.

Additional Simulated Beam Patterns



(a) (b) (c)

Figure A.1.: Layout of the icosahedral array (a), spherical harmonics subset (b) and mixed-order beam pattern simulated with the spherical cap model at f=200 Hz (c).



(a) (b) (c)

Figure A.2.: Layout of the dodecahedral array (a), spherical harmonics subset (b) and mixed-order beam pattern simulated with the spherical cap model at f=200 Hz (c).

Appendix B.

Directivity Plots of IKO1 and IKO3

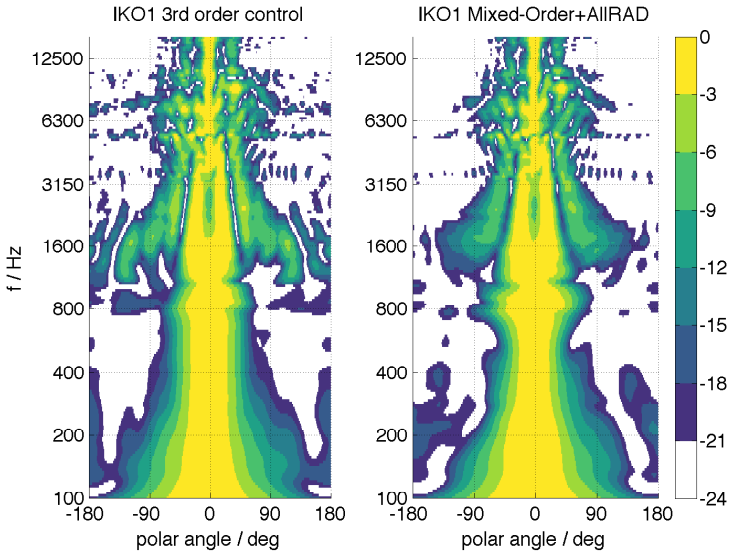


Figure B.1.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency.

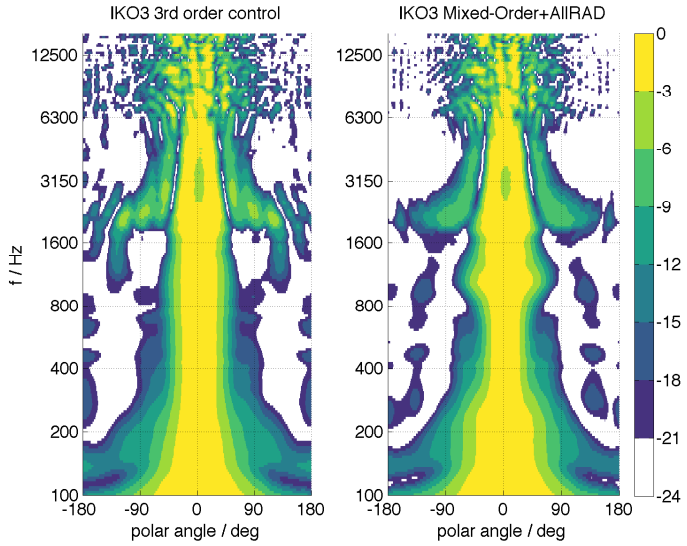


Figure B.2.: Horizontal cut of horizontal beams $\varphi = 0°$, $\theta = 90°$. Colorbar indicates relative dB, normalized to 0 dB for every frequency.