

Diploma Thesis

Modal beamforming using planar circular microphone arrays

Markus Zaunschirm

Graz, March 2012

Supervisor: Dr. Franz Zotter
Assessor: Prof. Robert Höldrich

Institute of Electronic Music and Acoustics

University of Music and Performing Arts Graz, Austria
Graz University of Technology, Austria



Abstract

Beamforming denotes the discrimination between signals based on the spatial location of sources. Whilst conventional beamforming algorithms directly operate on the sensor signals, modal beamforming approaches use directional modes that are obtained by decomposing the wave-field into orthogonal solutions of the acoustic wave equation. Spherical arrays are more generic as they allow for higher-order resolution and for steering a designed beampattern to any direction in the three-dimensional space without altering the shape. For applications where a full spherical array is uneconomic or redundant the use of circular arrays is feasible. As conventional circular arrays lack control of the beampattern in the vertical array plane, this thesis deals with the design of a planar circular microphone array that allows for steering beams towards both the horizontal and vertical array plane (but with changed shape). In this context, the decomposition into the directional modes is explained based on measurement data and based on an analytical model. Further, the general concept of modal beamforming is outlined and the generation of optimized beampatterns is derived based on linear pressure gradient microphone arrays.

Kurzfassung

Als Beamforming bezeichnet man die Unterscheidung von Signalen aufgrund ihres räumlichen Ursprungs. Während konventionelle Beamforming-Algorithmen direkt auf die Sensorsignale angewandt werden, verwenden modale Beamforming-Ansätze gerichtete und ausrichtbare Moden die durch die Zerlegung des Schallfeldes in orthogonale Funktionen gewonnen werden. Für gewöhnlich werden dazu kugelförmige Array-Geometrien verwendet, da sie die Ausrichtung eines gewünschten Richtmusters in jede beliebige Richtung im dreidimensionalen Raum ermöglichen. Für Anwendungen bei denen ein kugelförmiges Array redundant oder unökonomisch ist werden kreisförmige Arrays bevorzugt. Diese Arbeit behandelt das Design eines kreisförmigen Arrays mit dem es sowohl möglich ist eine gewünschte Richtwirkung ausrichtungsunabhängig in der horizontalen Ebene des Arrays zu erzeugen als auch die Empfindlichkeit gegenüber Quellen aus der vertikalen Ebene zu steuern. In diesem Zusammenhang wird die Zerlegung des Schallfelds in die orthogonalen Moden analytisch und basierend auf Messdaten beschrieben. Weiters wird das Konzept des modalen Beamformings erklärt und die Optimierung von Richtmustern bestimmter Ordnung wird anhand von linearen Druckgradienten-Arrays abgehandelt.

Acknowledgements

Special thanks go to my advisor, Dr. Franz Zotter, for his support and guidance during this work and my entire studies. His expert knowledge, flexible working hours and enthusiasm have been very helpful and motivating. I want to thank Hannes Pomberger.

Further, I want to thank my friends and fellow students for a terrific time.

This work is dedicated to my parents Elisabeth and Hans-Peter and the rest of my family, who supported me throughout the years of education and encouraged me to study what I was interested in.

Markus Zaunschirm
Graz, Austria, March 2011

Contents

1	Introduction	1
1.1	Thesis layout	6
2	Design of directivity patterns	8
2.1	For one direction with differential microphone arrays	8
2.1.1	Design using orthogonal Legendre polynomials	13
2.1.2	Design with constant side lobe attenuation	17
2.2	Variable directivity for adjustable directions (spherical microphone arrays)	21
2.2.1	Continuous spherical aperture	22
2.2.2	Re-normalization of the order weights	24
2.2.3	Discrete spherical arrays	24
2.2.4	Relation to hemispherical microphone arrays	26
3	Measuring the characteristics of the prototype	31
3.1	Measurement setup	32
3.1.1	Compensation of the transducer frequency response	34
3.1.2	Measured impulse responses	36
3.1.3	Time realignment	36
3.2	Directivity patterns	39
3.2.1	Displaying at discrete directions	39
3.2.2	Interpolated directivity patterns	40
3.2.3	Slices of the three dimensional directivity patterns	43
4	Modal SH beamforming based on measurements	46

4.1	Decomposition based on measurements	47
4.1.1	Practical inversion	52
4.1.2	Summary	58
4.2	Beamforming	58
4.2.1	White noise gain (WNG)	59
4.3	Implementation and practical aspects	62
4.3.1	Generated beampatterns	64
5	Analytic model of circular microphone arrays	69
5.1	Continuous circular aperture	69
5.2	Circular arrays	71
5.2.1	Sampling the circle	71
5.2.2	Frequency dependence of modes	72
5.2.3	Modal aliasing	74
5.2.4	Accessing the spherical harmonic modes	74
5.2.5	Higher order arrays	79
6	Compare model with measurement	81
6.1	Analysing the microphone directivity patterns	85
6.2	Holographic error	87
7	Summary and future work	89
7.1	Future work	90
A	Sound field description in spherical coordinates	91
A.1	Sound field theory	91
A.2	Solution of the Helmholtz equation	93
A.3	Spherical harmonics	96
A.4	Spherical harmonics transform	98
A.5	Discrete spherical harmonics transform	100
A.5.1	Matrix and vector notation	101
A.5.2	Spatial aliasing and sampling schemes	104
B	Calculation of order weights	107
B.1	Maximizing the DI	107

<i>CONTENTS</i>	vi
B.2 Maximizing front-to-back centroid	108
Bibliography	109

1. Introduction

Signal reception using microphone arrays offer a challenging but promising improvement or progression for a wide range of applications including signal enhancement, 3D soundfield recordings [BDM06], auditory scene analysis [Teu06], room acoustic measurements [BR04], plane-wave decomposition [DLZ⁺05] and modal beamforming [YSS⁺11] which is considered as the main focus of attention in this thesis.

Generally, signals of different sound sources interfere with each other if they occur at the same time and share the same frequency band. Accordingly, a separation by temporal filtering is not possible. But if the interfering signals originate from different locations in space it is possible to separate them by using a spatial filter. The terms spatial filtering and beamforming are closely related since early spatial filters were designed to form sharp beams that are sensitive to signals radiating from a specific direction while attenuating signals from other directions [VVB88] [Abh99]. Accordingly, beamforming describes the process of discriminating between signals based on the spatial source locations.

The implementation of a spatial filter requires a processing of data collected using an array of spaced transducers [Hay85] (cf. the application of a temporal filter requires the collection of data over a temporal aperture).

There are various designs for microphone arrays but geometrically most of those arrays are either linear, circular or spherical. The ideal design depends on the targeted application.

In short, beamforming is used to capture signals from directions of interest while signals from other directions are attenuated. Nevertheless, this is also partly achieved by directional acoustic transducers. Especially, unidirectional first order gradient mi-

crophones that exhibit characteristics specified by $(c + \cos(\vartheta))$, where c is a constant and ϑ is the polar angle between the main axis of the microphone and the direction of arrival of the incoming sound (see figure 1.1) are widely used. The term first order refers to the fact that for example a cardioid microphone is effectively a superposition of an omnidirectional and a figure-of-eight directional pattern. Additionally a

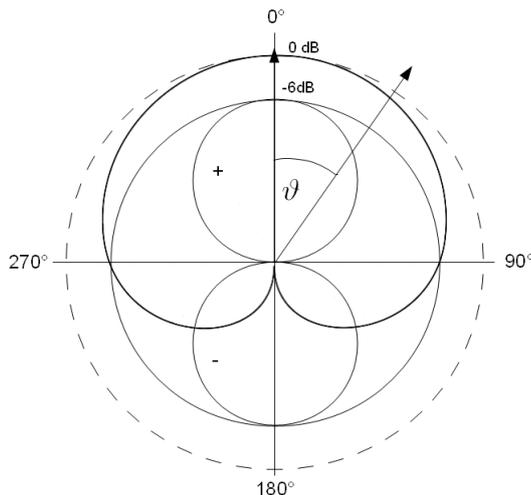


Figure 1.1: Polar plot of a unidirectional microphone with first order cardioid characteristics as a superposition of a omnidirectional and a first-order gradient microphone cf. [DfR08] fig. 4/20. The angle ϑ defines the polar angle between the main axis and the source position.

beamformer or a directive microphone can be characterized by its pick-up pattern. This directional pick-up pattern describes the output response of a beamformer and is defined as a function of space variables. Note that we use the term directional pick-up pattern interchangeable with the term beampattern. Figure 1.2 shows as an example the pick-up pattern of a first order cardioid microphone or respectively of a first order beamformer. It can be seen that the beampattern consists of a wide main lobe, no side lobes and a zero for $\vartheta = \pm\pi$. Accordingly, the distinction between the signal impinging from the main direction and interfering signals from other directions is not satisfying.

There are various ways of improving the directivity. In [SW88] it is stated that characteristics improve by utilizing second order gradient microphones. This is done by flush-mounting 2 first-order gradient microphones on a baffle.

However, the use of a single unidirectional microphone is restricted and lacks flexibility

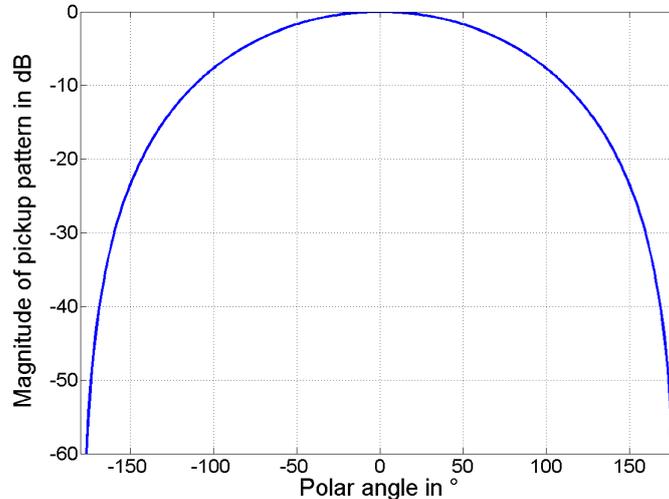


Figure 1.2: Example of a first order beampattern that is steered towards $\vartheta = 0$.

when compared to a microphone beamformer. This is because a beamformer can use the signals of a fixed microphone array to create various beampatterns, even such, having an adjustable direction and shape.

Most beamforming approaches process the sensor signals immediately and generally, require assumptions about the band width of the recorded signal. In particular, the array size determines the lower frequency bound and the separation of the array microphones determines an upper frequency bound.

In order to steer the beam direction, the simplest method, the delay-and-sum beamformer delays the sensor signals before summing them up. However, this beamforming approach does not achieve a frequency independent beam width. Therefore, filter-and-sum beamforming techniques are frequently employed to stabilize the beam width. For more detailed information about conventional array processing techniques the reader is referred to the works of [BW01] and [BCH08], which both offer an in depth review of this issue.

In contrast to conventional array processing algorithms modal array processing as applied in [Abh99], [ME02], [Teu06] uses signals transformed into domains of directional modes that are solutions of the acoustic wave equation. The solutions of the Helmholtz equation that apply to an interior problem in the spherical coordinate system are

$$j_n(kr)Y_n^m. \quad (1.1)$$

In eq. (1.1) the radial $j_n(kr)$ and the angular function Y_n^m appear clearly separated. Thus, the modes considered by modal beamforming are the functions Y_n^m , which are known as spherical harmonics. Arbitrary beampatterns can be represented by a linear combination of those spherical harmonics, see A.4.

One of the first modal approaches was presented by Gerzon and Graven in [CG77], where the well know sound-field microphone is established. The described assembly is aimed to provide output signals which would be obtained from a multitude of coincident ¹ microphones. In order to do so the raw audio data (A-format) recorded by four directional microphones (cardioid or hypercardioid) that are arranged in a tetrahedral layout, are mapped by a matrix multiplication to the so called B-format that contains the channels W,X,Y and Z. The directional characteristics of those channels are related to the spherical harmonic signals up to the order one [Rei06], where W contains the signal of an omnidirectional transducer and the channels X,Y and Z correspond to signals picked up with figure-of-eight transducers directed towards the three Cartesian coordinate directions. Combining these signals is used to model any first order beampattern direction and shape and it is possible to adapt the beampattern in a desired manner in realtime. Altogether, the four signals contain a first-order resolution surround sound image.

Rigid spherical arrays are more general as they permit sound-field capture with higher-order resolution. However, a full spherical array may be uneconomic or redundant in real-world scenarios where acoustic sources are restrained to the upper half of the three-dimensional space. For example Li and Duraiswami [LR05] designed a hemispherical table microphone for sound capture and beamforming. This approach is legitimate if all sources are assumed to be above a table or other boundary surfaces, but due to the actual numerousness of microphones, its artless usage and industrial producibility has to be considered rather difficult and problematic.

The company AKG already holds the patent on a planar circular table microphone where the microphones are positioned coincident on the array and the decomposition of the acoustic scene in the previously mentioned modes is done up to the first order [Rei09]. The existing array consist of three cardioid microphones, that are ideally

¹Coincident means that the microphones are positioned at the same location and accordingly there is no transmission time difference between the microphones. Physically it is not possible to have two or more separate microphones at precisely the same position.

located at one central point (see figure 1.3(a)). In order to obtain spherical or circular

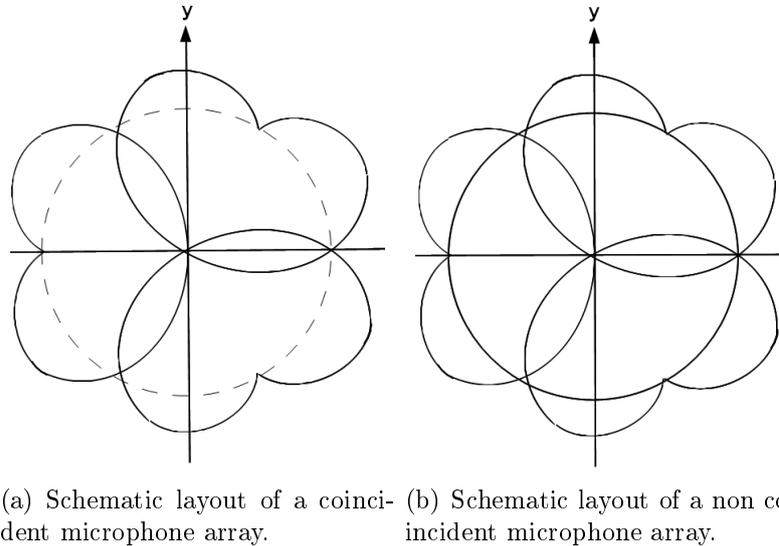


Figure 1.3: Schematic layouts of planar circular microphone arrays.

harmonic signals out of the raw microphone data it is necessary to combine and weight the signals of the cardioid microphones (cf. soundfield microphone). As just the even-symmetric spherical harmonics with respect to the horizontal plane satisfy the boundary condition, the skew symmetric harmonics are not considered. In the first order case this means that the upward-pointing figure-of-eight spherical harmonic is omitted, see fig. 1.4(a). Therefore, it is not possible to control and steer beams in the vertical direction of the array.

By placing the three cardioid microphones on a concentric circle and adding an omnidirectional microphone in the centre of the array (see fig. 1.3(b)) it is possible to access an additional spherical-harmonic that also enables controlling the vertical directivity of the array, see fig. 1.4(b).

In order to further improve the order of the possible pick-up patterns, the work presented in this thesis deals with the challenging design of a comparable planar array that is capable of achieving the next higher spatial resolution. A possible solution to this problem is outlined in literature by Meyer and Elko in [ME08a] and the patent application [ME08b]. This approach is quite promising but still leaves room for improvements.

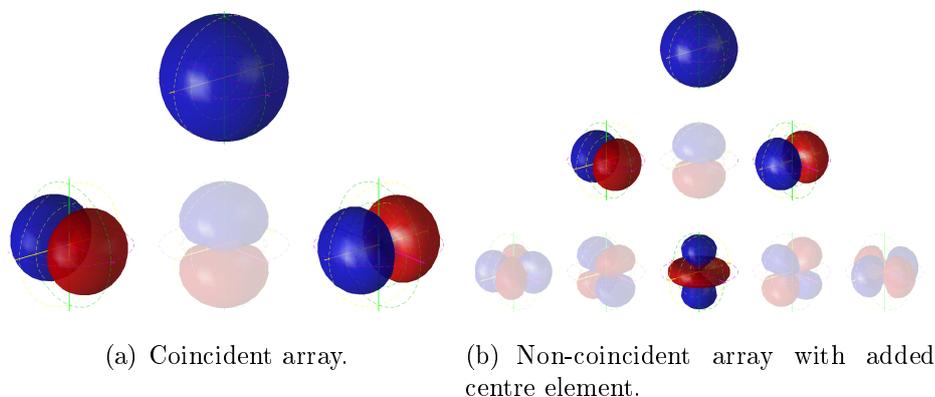


Figure 1.4: Accessible spherical harmonics of the mentioned arrays.

1.1 Thesis layout

Chapter 2 explains the design of optimized pick-up patterns with pressure gradient microphone arrays as well as the concept of modal beamforming using spherical microphone arrays. The general pick-up patterns are expressed in terms of polynomials and optimized using the orthogonal Legendre polynomials. A transition from the linear differential array to spherical microphone arrays is found via the spherical harmonics addition theorem and thus the optimized weightings in order to achieve a specific beam pattern are also applicable to modal spherical harmonics beamforming. Finally, idealized pick-up patterns and the corresponding weightings to achieve them are presented.

The necessary steps to measure and display frequency dependent three-dimensional directivity patterns of the individual array sensors are explained on the example of a circular planar microphone array in chapter 3. Further, spherical harmonics interpolation is used to evaluate the directivity patterns on a smoother grid. Slices through the patterns are presented in order to analyse the rotational symmetry.

Chapter 4 presents a modal beamforming approach for the circular array prototype that is based on measurement data. Starting from a MIMO-system description of the measurement data in the spherical harmonics domain, filters that decompose the microphone signals into signals that are finally used for modal beamforming is discussed. A real-time implementation in pure data (pd) and the practically achievable

beampatterns are presented.

The analytic model of modal beamforming with planar circular microphone arrays is presented in chapter 5. Particular attention is paid to the concept of modal aliasing and the separation of aliased modes that is necessary to gain control over the vertical response of the planar array. Further, the design of higher order circular arrays is outlined.

Chapter 6 provides a comparison of the analytic model and the measured system. Occurring mismatches and the influence of array imperfections are analysed.

2. Design of directivity patterns

2.1 For one direction with differential microphone arrays

The design and fundamentals of differential microphone arrays is briefly revised based on [HB04] in this section. Note that all presented derivations assume acoustic plane wave propagation and thus let us define the acoustic pressure thereof as

$$p(\mathbf{k}, \mathbf{z}, t) = p_0 e^{i(\omega t + \mathbf{k}^T \mathbf{z})} = p_0 e^{i(\omega t + kz \cos \vartheta)}, \quad (2.1)$$

where p_0 is the plane-wave amplitude, ω is the angular frequency and T is the transpose operator. The acoustic-wave vector \mathbf{k} depends on the azimuth (φ) and zenith angle (ϑ) and defines the direction of arrival of the plane wave

$$\mathbf{k} = k \boldsymbol{\theta} = k \begin{pmatrix} \cos(\varphi) \sin(\vartheta) \\ \sin(\varphi) \sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix}, \quad (2.2)$$

where $k = \|\mathbf{k}\| = \omega/c$. In a same manner the orientation or position of the linear microphone array is defined by the vector $\mathbf{z} = (0, 0, z)^T$ and thus, the angle between the position vector and the wave-vector \mathbf{k} is defined as the polar angle ϑ^1 . By dropping the time dependence the n^{th} order spatial derivative along the direction \mathbf{z} can be written as

$$\frac{\partial^n}{\partial z^n} p(k, z, \vartheta) = p_0 (ik \cos \vartheta)^N e^{+ikz \cos \vartheta}, \quad (2.3)$$

¹The scalar product of two vectors is defined as the product of the vector lengths and the cosine of the angle between the vectors.

and thus it can be concluded from eq. (2.3) that an n^{th} order differential exhibits a bidirectional pattern with a shape defined by $(\cos \vartheta)^n$. Further, it is easy to see that the according frequency response is a high-pass with a slope of $6n$ dB per octave.

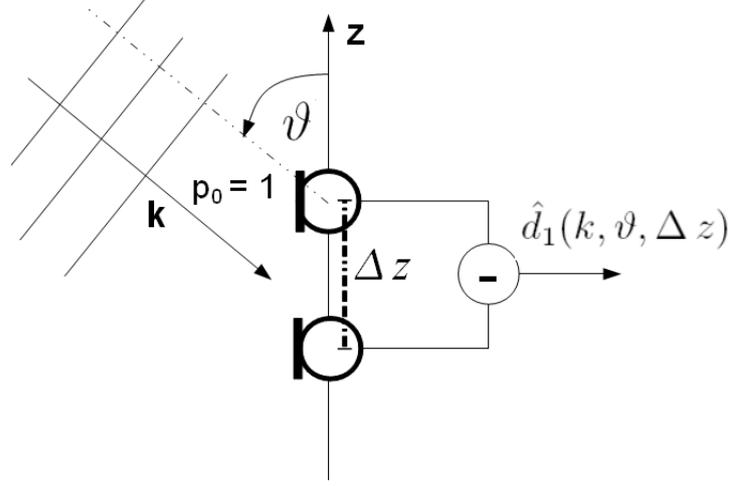


Figure 2.1: Scheme of first order difference sensor composed of 2 omnidirectional (zero-order) microphones.

The sampling of the acoustic pressure along the observation direction as depicted in fig 2.1 results in a finite-difference approximation of the spatial derivative where the first order approximation can be written as

$$\lim_{\Delta z \rightarrow 0} \frac{\Delta p(k, z, \vartheta)}{\Delta z} = \lim_{\Delta z \rightarrow 0} \frac{p(k, z + \Delta z/2, \vartheta) - p(k, z - \Delta z/2, \vartheta)}{\Delta z}, \quad (2.4)$$

where Δz denotes the microphone spacing and the $\lim_{\Delta z \rightarrow 0}$ indicates that the microphone spacing needs to be really small. Meyer [HB04] defines an amplitude bias that is the ratio of the analytic derivative to the approximated difference of a first order system and concludes that the element spacing of the two closely-spaced omnidirectional microphones should be less than $1/4$ of the acoustic wavelength in order to achieve an error that is less than 1 dB.

Now let us again consider the simple configuration shown in fig. 2.1. For a plane-wave as defined by eq. (2.1) the output of this two element sensor with interelement spacing

Δz is obtained from

$$\begin{aligned}\hat{d}_1(k, \vartheta, \Delta z) &= e^{ik\frac{\Delta z}{2} \cos \vartheta} - e^{-ik\frac{\Delta z}{2} \cos \vartheta} \\ &= 2i \sin(k\Delta z/2 \cos \vartheta),\end{aligned}\tag{2.5}$$

where the subscript indicates the first order differential and the arguments denote the dependence on the wavenumber, the element spacing and the polar angle ϑ . With a simplification ² the sensor output is rewritten as

$$\hat{d}_1(k, \vartheta, \Delta z) \approx ik\Delta z \cos \vartheta,\tag{2.6}$$

where the factor $\cos \vartheta$ again describes the beam pattern shape. Further, it is noticed from eq. (2.6) that the first order differential exhibits a high pass frequency response.

The generation of higher order differentials can be done as depicted in fig. 2.2 by extending the simple case of a first order sensor depicted in fig. 2.1. In words this means that a sensor that provides a measure of the n^{th} derivative of the incident pressure field is composed of $m p^{\text{th}}$ order microphones where $m + p - 1 = n$ [HB04]. After a frequency response compensation of each order with $1/(ik\Delta z)^n$ one obtains the normalized, frequency independent differentials

$$d_n(\vartheta) = \frac{\hat{d}_n(k, \vartheta, \Delta z)}{(ik\Delta z)^n}\tag{2.7}$$

that are shown in fig. 2.3.

² $\sin(k\Delta z/2 \cos \vartheta) \approx k\Delta z/2 \cos \vartheta$

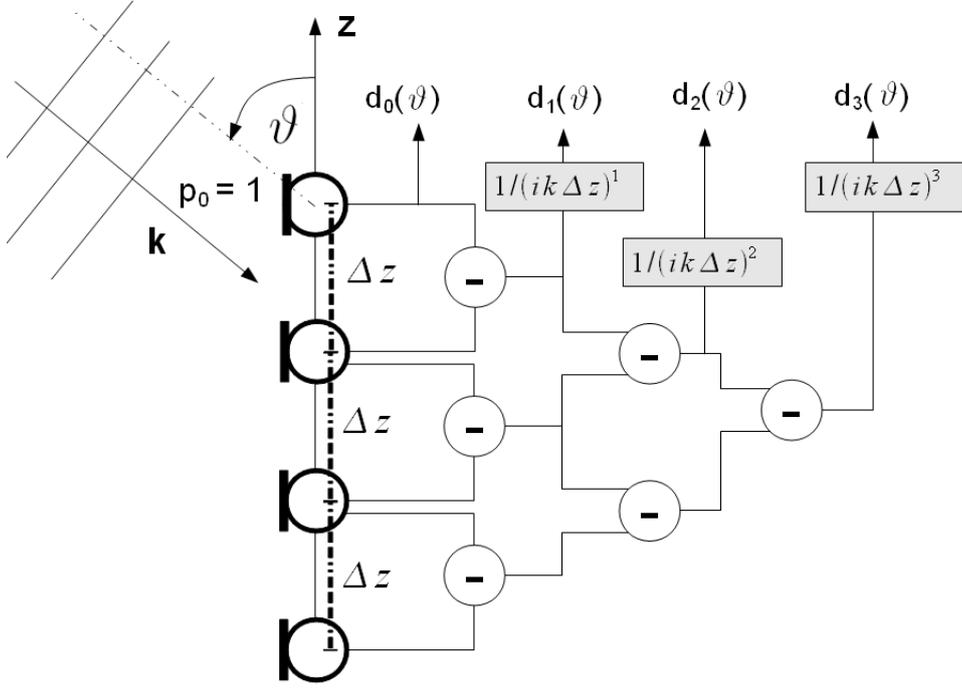


Figure 2.2: Generation of higher order differential sensor as a combination of lower order differentials.

Combining order outputs: Once the normalized differential outputs are available they can be weighted and combined in order to form specific array beampatterns of the maximum order N

$$\begin{aligned}
 d(\vartheta) &= a_0 + a_1 \cos \vartheta + a_2 \cos^2 \vartheta \dots + a_N \cos^N \vartheta \\
 &= \sum_{n=0}^N a_n \cos^n \vartheta,
 \end{aligned} \tag{2.8}$$

where the sum of the weighting coefficients should follow the rule $\sum_{n=0}^N a_n = 1$ in order to obtain a main axis steered towards $\vartheta = 0^\circ$ with $d(0^\circ) = 1$.

As a beamforming microphone array can be used in various environments or in general in different situations, also the requirements can change regarding the positioning of the zeros of the beampattern or the directivity gain towards the look direction. Therefore the weights of the different order differentials can be determined based on specific optimization criteria as the directivity index DI or the front-to-back ratio. And in order to suppress interfering signals from specific directions it is feasible to

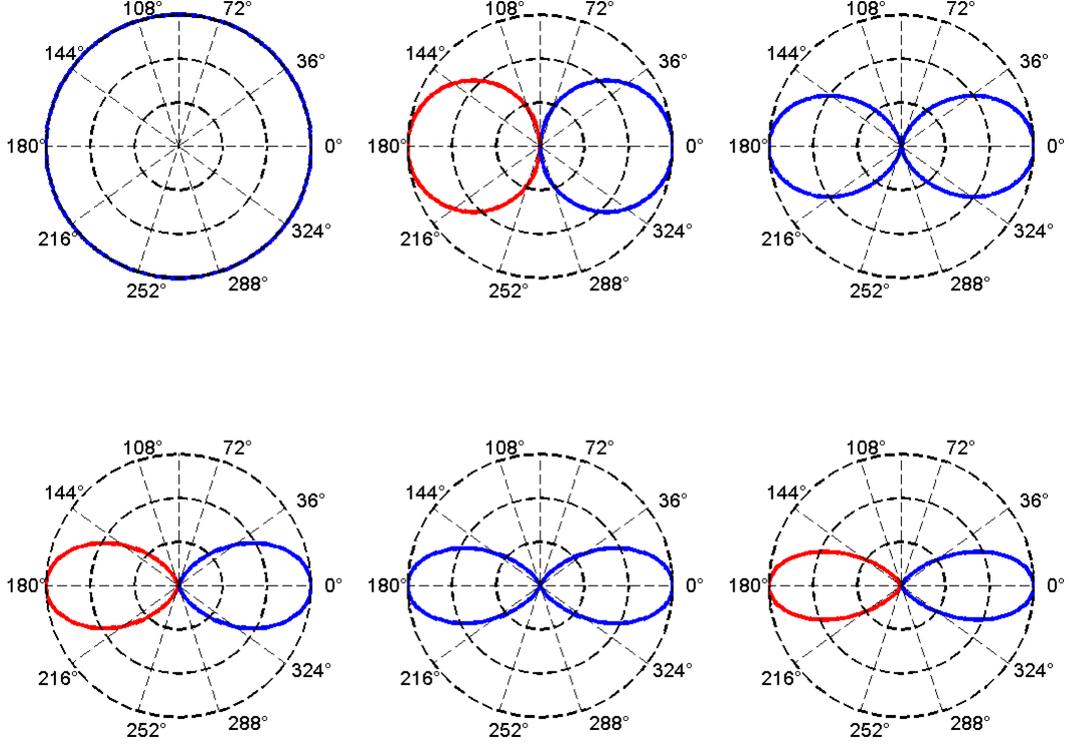


Figure 2.3: Pick-up pattern (10dB/division) of equalised differentials with $n = 0, \dots, 5$.

design the zeros or the minima of the directivity pattern for these directions.

In the following the directivity patterns are expressed in terms of polynomials $\mathcal{P}_n(\mu)$ where $\mu = \cos \vartheta$. The differential signal weights can then be found by comparing coefficients with the N^{th} order polynomial defined in eq. (2.8).

Cardioid directivity: A cardioid pattern is designed to suppress signals impinging from $\vartheta = 180^\circ$. The design criterion of an cardioid directivity response is that it exhibits a maximally flat response around that zero [Dan00]. A polynomial that combines those two properties and describes a cardioid directivity pattern of the order N is defined as

$$d(\mu) = \frac{(\mu + 1)^N}{2^N}. \quad (2.9)$$

The maximal flatness criterion is defined as

$$\frac{\partial^n d(\mu)}{\partial \mu^n} \Big|_{\mu=-1} = 0 \quad \text{for } n = 1, \dots, N-1 \quad (2.10)$$

which means that all derivatives of the cardioid directivity pattern up to order $N-1$ are zero at $\mu = -1$. And thus, the cardioid pattern has an $(N-1)$ -fold zero over the whole range of definition at $\mu = -1$ and changes its value as slowly as possible as we move from that point. The according weightings a_n can be obtained by expanding the polynomial defined in eq. (2.9) to the general form defined by eq. (2.8). Note that there is a relation to Butterworth filters since they are also designed to exhibit a maximally flat frequency response in the passband.

Compactly, the differential coefficients in order to achieve a cardioid pick-up pattern of order N are calculated by the binomial coefficients

$$a_n = \frac{\binom{N}{n}}{2^N}. \quad (2.11)$$

The corresponding first order and second order cardioid characteristics are exemplarily

$$d(\vartheta) = \frac{1 + \cos \vartheta}{2} \longrightarrow a_0 = 1/2; a_1 = 1/2, \text{ for } N = 1 \quad (2.12)$$

$$d(\vartheta) = \frac{1 + 2 \cos^2 \vartheta + \cos \vartheta}{4} \longrightarrow a_0 = 1/4; a_1 = 1/2; a_2 = 1/4, \text{ for } N = 2. \quad (2.13)$$

2.1.1 Design using orthogonal Legendre polynomials

For further optimization considerations we use the Legendre polynomials to express the directivity patterns. The Legendre polynomials form a polynomial sequence of orthogonal polynomials over the interval of $[-1, 1]$

$$\int_{-1}^1 P_n(\mu) P_{n'}(\mu) d\mu = \frac{2\delta_{n,n'}}{2n+1} \quad (2.14)$$

with the property $P_n(1) = 1$, where n determines the degree of the Legendre polynomial [JJ00].

Accordingly, a function of finite degree can be expressed as a superposition of weighted Legendre polynomials

$$d(\mu) = \sum_{n=0}^N \hat{w}_n P_n(\mu). \quad (2.15)$$

As all polynomials, the Legendre polynomials of order n only contain powers $\mu^{n'}$ up to $n' \leq n$ it is possible to express them as sum of its polynomial coefficients

$$P_n(\mu) = \sum_{n'=0}^n p_{n,n'} \mu^{n'}, \quad (2.16)$$

where $p_{0,0} = p_{1,1} = 1$ and $p_{n,n'} = 0$ for odd $n + n'$ and for $n' > n$. The coefficients are obtained using the three term recurrence relation

$$p_{n+1,n'} = \frac{2n+1}{n+1} p_{n,n'-1} - \frac{n}{n+1} p_{n-1,n'} \quad (2.17)$$

see table 2.1.

order $n =$	$P_n(\mu) = \sum_{n'=0}^n p_{n,n'} \mu^{n'}$									
	$p_{n,9}$	$p_{n,8}$	$p_{n,7}$	$p_{n,6}$	$p_{n,5}$	$p_{n,4}$	$p_{n,3}$	$p_{n,2}$	$p_{n,1}$	$p_{n,0}$
0										1
1									1	
2								$\frac{3}{2}$		$-\frac{1}{2}$
3							$\frac{5}{2}$		$-\frac{3}{2}$	
4						$\frac{35}{8}$		$-\frac{30}{8}$		$\frac{3}{8}$
5					$\frac{63}{8}$		$-\frac{70}{8}$		$\frac{15}{8}$	
6				$\frac{231}{16}$		$-\frac{315}{16}$		$\frac{105}{16}$		$-\frac{5}{16}$
7			$\frac{429}{16}$		$-\frac{693}{16}$		$\frac{315}{16}$		$-\frac{35}{16}$	
8		$\frac{6435}{128}$		$-\frac{12012}{128}$		$\frac{6930}{128}$		$-\frac{1260}{128}$		$\frac{35}{128}$
9	$\frac{12155}{128}$		$-\frac{25740}{128}$		$\frac{18018}{128}$		$-\frac{4620}{128}$		$\frac{315}{128}$	

Table 2.1: Legendre polynomials with the variable argument $\mu = \cos \vartheta$.

Thus, it is feasible to define a matrix \mathbf{P} that contains the coefficients of the Legendre polynomials of order n in column n

$$\mathbf{P} = \begin{pmatrix} p_{0,0} & 0 & p_{2,0} & 0 & \dots \\ 0 & p_{1,1} & 0 & p_{3,1} & \dots \\ \vdots & 0 & p_{2,2} & 0 & \dots \\ \vdots & \vdots & 0 & p_{3,3} & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix} \quad (2.18)$$

The weights of the differential signals are then determined by a matrix multiplication

$$\mathbf{a} = \mathbf{P} \hat{\mathbf{w}} \quad (2.19)$$

where $\mathbf{a} = [a_0, \dots, a_N]^T$ contains the weights of the corresponding signals μ^n and $\hat{\mathbf{w}} = [\hat{w}_0, \dots, \hat{w}_N]^T$ is a vector containing the order weights of the Legendre polynomials.

Note that Chebyshev polynomials would also be orthogonal on the interval $[-1, 1]$ but with respect to a weight $\frac{1}{\sqrt{1-\mu^2}}$. Thus, using the orthogonal Legendre polynomials for optimizing measures that include the squared pick-up pattern is preferred since $\int_{-1}^1 d(\mu)^2 d\mu$ equals $\sum_{n=0}^N \hat{w}_n^2 \frac{2}{2n+1}$ when expressed in Legendre polynomials.

Hypercardioid characteristics: Hypercardioid microphones are designed to achieve maximal directivity, see eq. (2.45). Consequently, the optimal weights are found by maximizing the simplified directivity factor

$$D = \frac{\int_{-1}^1 1^2 d\mu}{\int_{-1}^1 d(\mu)^2 d\mu} \quad (2.20)$$

that compares the squared averaged directional sensitivity of a directional microphone with an omnidirectional microphone. It is obvious from eq. (2.20) that the denominator needs to be minimized under the constraint of $d(1) = 1$ in order to maximize the directivity factor.

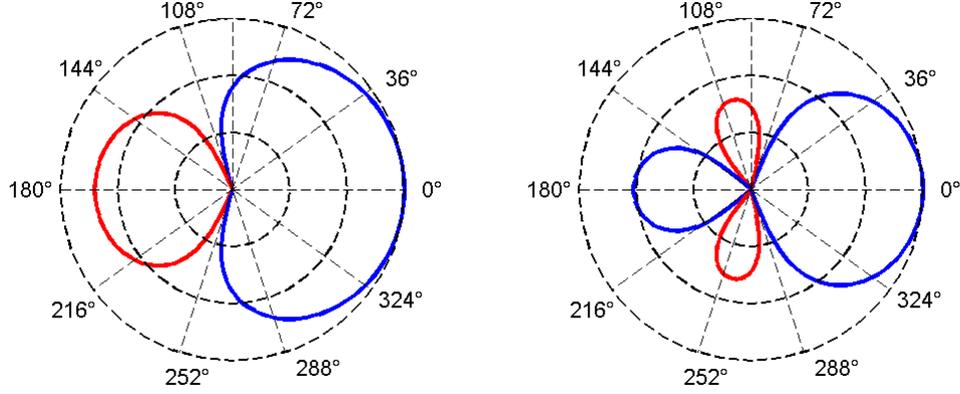
This constrained optimization problem is solved using a Lagrange multiplier and the according optimal weights are found to be

$$\hat{w}_n = \frac{2n+1}{(N+1)^2}, \quad (2.21)$$

where \hat{w}_n are the weighting coefficients, see appendix B. Optimal weights for the differential signals are determined according to eq. (2.19).

Figure 2.4 shows the pick-up patterns of a first and second order hypercardioid pick-up pattern.

Supercardioid characteristics: Another possible optimization criterion is based on maximizing the front-to-back ratio (or front-to-rear ratio) that is defined as the ratio of the squared directional gain between the front and the rear of the directivity pattern [HB04]. The same definition is given in [Bau42], where it is stated that a unidirectional microphone should accept sound arriving from a hemisphere at the



(a) 1st order hypercardioid pick-up pattern. (b) 2nd order hypercardioid pick-up pattern.

Figure 2.4: Pick-up patterns (10dB/division) with maximal directivity index.

front and reject sound arriving from a hemisphere at the back. The optimum pick-up pattern in accordance with this definition is termed supercardioid.

However, throughout this thesis a supercardioid weighting \hat{w}_n is obtained by maximizing the front-to-back centroid of a beampattern with highest sensitivity for $\mu = 1$.

Let us define the front-to-back centroid of a beampattern $d(\mu)$ as

$$r_E = \frac{\int_{-1}^1 \mu d(\mu)^2 d\mu}{\int_{-1}^1 d(\mu)^2 d\mu}. \quad (2.22)$$

With the recursion formula of Legendre polynomials [MI86]

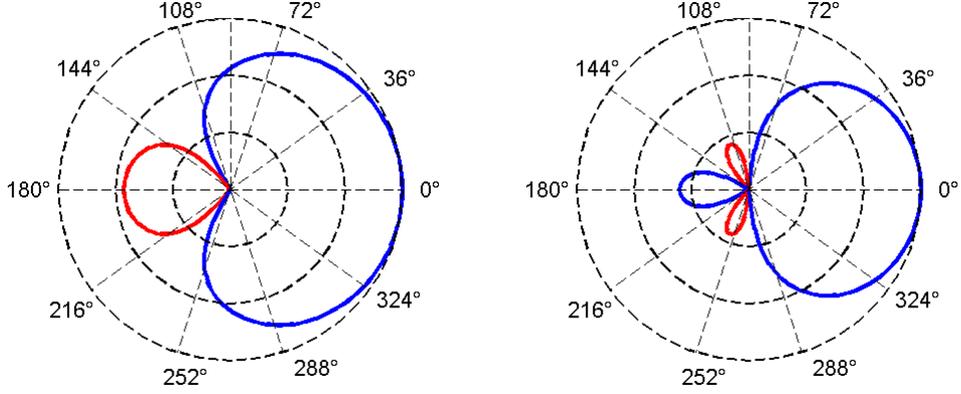
$$(n+1)P_{n+1}(\mu) = (2n+1)\mu P_n(\mu) - nP_{n-1}(\mu), \quad (2.23)$$

and their orthogonality the approximated optimal solution is found to be

$$\hat{w}_n = P_n \left(\cos \left(\frac{137.9^\circ}{N+1.52} \right) \right) (2n+1). \quad (2.24)$$

The relevant calculations are presented in the appendix B

Figure 2.5 shows the pick-up patterns of a first and second order supercardioid pick-up pattern.



(a) 1st order supercardioid pick-up pattern. (b) 2nd order supercardioid pick-up pattern.

Figure 2.5: Pick-up patterns (10dB/division) with maximal front-to-back centroid.

Cardioid characteristic with Legendre polynomials: In literature this optimization approach is often referred to as in-phase approach, [DRP98], [Dan00]. Let us again consider the beampattern $d(\mu) = \sum_{n=0}^N \hat{w}_n P_n(\mu)$. Then the weights \hat{w}_n can be found according to

$$\frac{\partial^n d(\mu)}{\partial \mu^n} \Big|_{\mu=-1} = 0, \quad \text{for } n = 0, \dots, N-1, \quad (2.25)$$

which again forces the directivity pattern to have the only zero at $\mu = -1$ and to be maximally flat around that zero [Dan00]. Since the necessary derivations in order to determine the corresponding order weightings are not generally straightforward we subject the determination to the previously established nested comparison of coefficients defined in eq. (2.19) (the analytic solution is presented in [Dan00]). Accordingly the order weights \hat{w}_n are determined by

$$\hat{\mathbf{w}} = \mathbf{P}^{-1} \mathbf{a}. \quad (2.26)$$

Figure 2.6 shows the pick-up patterns of a first and second order cardioid pick-up pattern.

2.1.2 Design with constant side lobe attenuation

Another possible approach is to design for constant side lobe attenuation, which is mentioned for the sake of completeness. In the same manner as Legendre polynomials

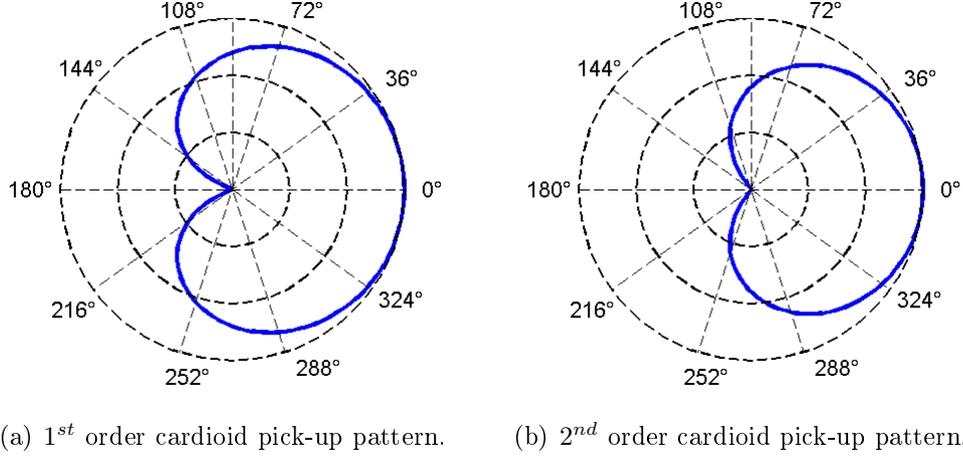


Figure 2.6: Pick-up patterns (10dB/division) with maximally flat zero.

are used to design appropriate beampatterns it is also possible to use the Chebyshev polynomials of the first kind which make from a cosine argument in the range of $[-1; 1]$ a cosine of n times its phase. However, it also works for arguments exceeding the range of $[-1; 1]$ when defined with cosine and arcus cosine for imaginary arguments. In general, evaluation of the Chebyshev polynomials yields

$$T_n(\tilde{\mu}) = \cos(n \arccos \tilde{\mu}). \quad (2.27)$$

As argument one can use $\tilde{\mu} = [(1 + \mu)\alpha - 1]$, where the factor α can be used to set the wanted side-lobe attenuation that can be determined order dependent as

$$\alpha = \frac{1}{2} \left(\frac{\cos(\arccos 10^{\tilde{\alpha}/20})}{n} + 1 \right) \quad (2.28)$$

where $\tilde{\alpha}$ is the wanted side lobe attenuation in dB . As all polynomials the Chebyshev polynomials of order n contain powers $\tilde{\mu}^{n'}$ up to $n' \leq n$ and accordingly, the differential signal weights a_n are obtained by comparison of coefficients of the expanded Chebyshev polynomial with the polynomial defined in eq. (2.8). Again, the polynomial coefficients $t_{n,n'}$ (see table 2.2) of the Chebyshev polynomial are obtained using a three term recurrence

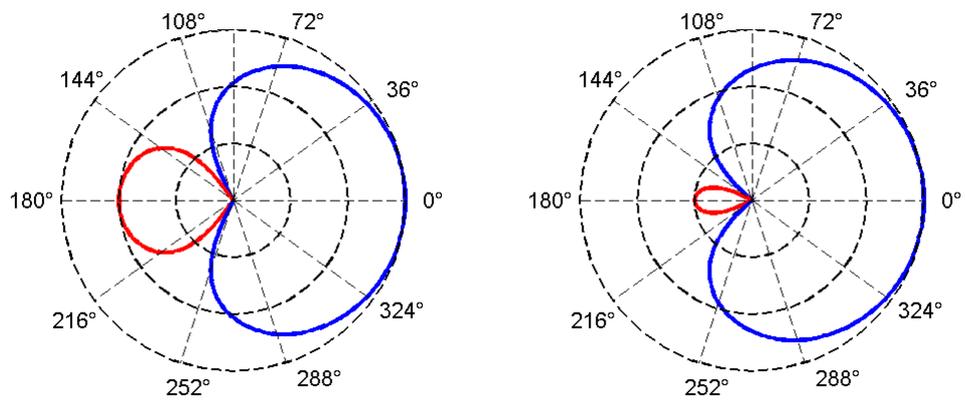
$$t_{n+1,n'} = 2t_{n,n'-1} - t_{n-1,n'} \quad (2.29)$$

where $t_{0,0} = t_{1,1} = 1$ and $t_{n,n'} = 0$ for odd $n + n'$ and for $n' > n$, c.f. Legendre polynomials.

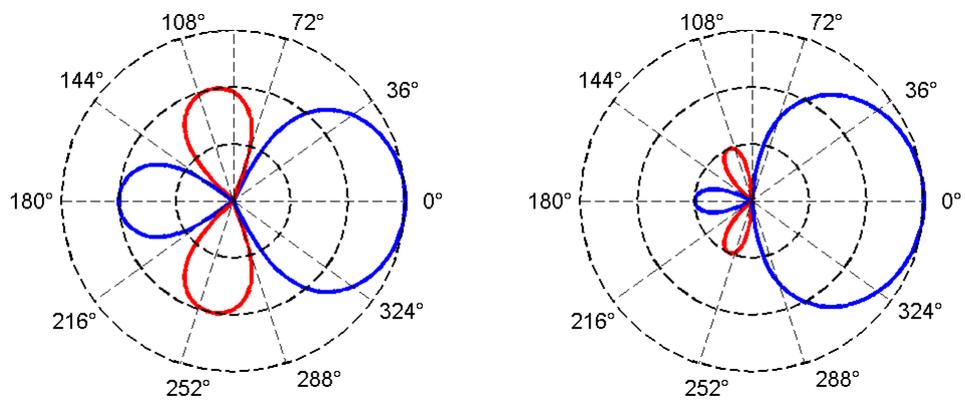
order $n =$	$T_n(\tilde{\mu}) = \sum_{n'=0}^n t_{n,n'}$									
	$t_{n,9}$	$t_{n,8}$	$t_{n,7}$	$t_{n,6}$	$t_{n,5}$	$t_{n,4}$	$t_{n,3}$	$t_{n,2}$	$t_{n,1}$	$t_{n,0}$
0										1
1									1	
2								2		-1
3							4		-3	
4						8		-8		1
5					16		-20		5	
6				32		-48		18		-1
7			64		-112		56		-7	
8		128		-256		160		-32		1
9	256		-576		432		-120		9	

Table 2.2: Chebyshev polynomials with the variable argument $\tilde{\mu} = [(1 + \mu) \alpha - 1]$.

The obtained directivity patterns with constant side lobe attenuation are depicted for first and second order in fig. 2.7.



(a) First order with side lobe attenuation of 10dB. (b) First order with side lobe attenuation of 20dB.



(c) Second order with side lobe attenuation of 10dB. (d) Second order with side lobe attenuation of 20dB.

Figure 2.7: Directivity patterns (10dB/division) with constant side lobe attenuation.

2.2 Variable directivity for adjustable directions (spherical microphone arrays)

As the quantity to be measured on the sphere can be the pressure or a combination of pressure and velocity the variable x is used instead of the pressure p to denote the quantity of interest.

Now, let us define the sound-field at an observation point $\boldsymbol{\theta}$ on a virtual spherical surface of radius r due to a plane wave impinging from the z -direction (see fig. 2.8(a)) ($\vartheta_i = 0$) in spherical coordinates and without time-dependency as

$$x(kr, \vartheta) = e^{i(kr \cos \vartheta)} = \sum_{n=0}^{\infty} (2n+1) i^n \rho_n(kr) P_n(\cos \vartheta), \quad (2.30)$$

where r and ϑ define the observation point and $\rho_n(kr)$ is a function which depends on the observation radius and the frequency of the plane wave that is propagating towards the origin of the coordinate system. For further reading and definitions the reader is referred to the appendix A or [Zot08] and [EKM04].

With the addition theorem [EKM04]

$$P_n(\langle \boldsymbol{\theta}_i, \boldsymbol{\theta} \rangle) = \sum_{m=-n}^n \frac{(n-m)!}{(n+m)!} P_n^m(\cos \vartheta_i) P_n^m(\cos \vartheta) e^{im(\varphi_i - \varphi)} \quad (2.31)$$

it is possible to find a general expression of the sound field at an observation point $\boldsymbol{r} = r \boldsymbol{\theta}$ due to an plane wave impinging from any direction $\boldsymbol{\theta}_i$ (see fig. 2.8(b)) as

$$x(kr, \boldsymbol{\theta}_i, \boldsymbol{\theta}) = 4\pi \sum_{n=0}^{\infty} i^n \rho_n(kr) \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}_i) Y_n^m(\boldsymbol{\theta}), \quad (2.32)$$

where Y_n^m are the real valued spherical harmonics of order n and degree m (see sec. A.3) and $\rho(kr)$ depends on the sphere configuration [BR07]

$$\rho_n(kr) = \begin{cases} j_n(kr) & \text{open-sphere omnidirectional characteristic} \\ (j_n(kr) - i j_n'(kr)) & \text{open-sphere cardioid characteristic} \\ \left(j_n(kr) - \frac{j_n'(kr)}{h_n'(kr)} h_n(kr) \right) & \text{rigid-sphere.} \end{cases} \quad (2.33)$$

Note that eq. (2.33) describes the sound field at the sphere surface.

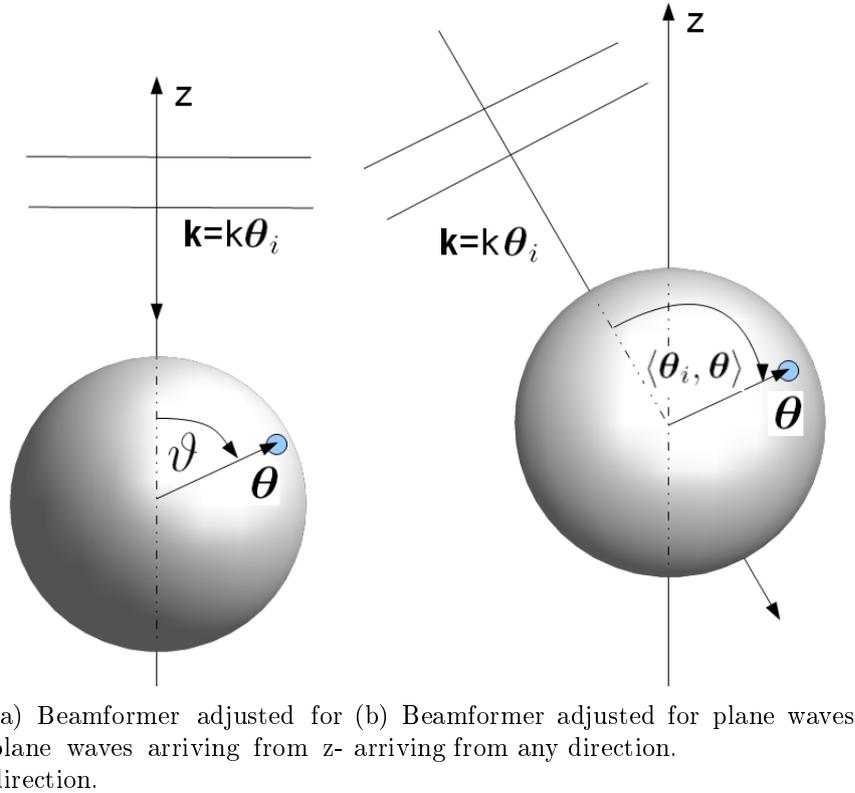


Figure 2.8: Scheme of a spherical aperture.

2.2.1 Continuous spherical aperture

The spherical harmonics transform (SHT) of the sound-field quantity x on the sphere is calculated by

$$\chi_n^m(kr, \boldsymbol{\theta}_i) = \int_{S^2} x(kr, \boldsymbol{\theta}, \boldsymbol{\theta}_i) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} = 4\pi i^n \rho_n(kr) Y_n^m(\boldsymbol{\theta}_i), \quad (2.34)$$

where $\chi_n^m(kr, \boldsymbol{\theta}_i)$ denotes the spherical wave spectrum coefficient of the sound-field at the sphere radius r . For further reading on the spherical harmonics transform the reader is referred to the appendix A.4.

Due to the fact that any square integrable function on the sphere can be expanded in terms of spherical harmonics, it is possible to describe any beampattern with this sort of functions. Let us define the expansion of an ideal beampattern looking at direction

$\boldsymbol{\theta}_1$ that can be modelled as a delta function

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \delta(1 - \boldsymbol{\theta}_i^T \boldsymbol{\theta}_1), \quad (2.35)$$

as

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}_i) Y_n^m(\boldsymbol{\theta}_1). \quad (2.36)$$

Now let us define an aperture weighting function $g(kr, \boldsymbol{\theta})$ such that multiplying it with the recorded sound-field and integrating the product over the entire sphere yields a frequency independent beampattern

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \int_{S^2} x(kr, \boldsymbol{\theta}, \boldsymbol{\theta}_i) g(kr, \boldsymbol{\theta}, \boldsymbol{\theta}_1) d\boldsymbol{\theta}. \quad (2.37)$$

By substituting eq. (2.34) and the weighting function in the spherical harmonics domain $g(kr, \boldsymbol{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n g_n^m(kr) Y_n^m(\boldsymbol{\theta})$ into eq. (2.37) the targeted beampattern is defined as

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \chi_n^m(kr, \boldsymbol{\theta}_i) g_n^m(kr, \boldsymbol{\theta}_1). \quad (2.38)$$

As the coefficients $\chi_n^m(kr, \boldsymbol{\theta}_i)$ are frequency dependent, the aperture weighting function needs to compensate for the dependency that is given by $\rho(kr)$. Consequently, the spherical harmonic coefficients of the weighting function, in order to obtain the frequency independent beampattern defined in eq. (2.36), are defined as

$$g_n^m(kr, \boldsymbol{\theta}_1) = \frac{1}{4\pi i^n \rho_n(kr)} Y_n^m(\boldsymbol{\theta}_1). \quad (2.39)$$

In order to design adjustable beampattern shapes, order weightings w_n are inserted into eq. (2.36) and thus, the designed array directivity in dependence of the look direction and the arrival direction of a plane wave $\boldsymbol{\theta}_i$ is computed by substituting the defined aperture weighting coefficients (2.39) and the sound-field coefficients (2.34) into eq.(2.38)

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}_i) w_n Y_n^m(\boldsymbol{\theta}_1). \quad (2.40)$$

Note that the order weightings w_n are designed to optimize the beampatterns for different applications, see section 2.

The ideal beam can be steered into any direction $\boldsymbol{\theta}_1$. But this is only true for the ideal case of an ideal continuous spherical aperture and for infinite summations.

2.2.2 Re-normalization of the order weights

Since the modal spherical harmonics beamformer uses the energy normalized Legendre polynomials the order weights \hat{w}_n derived for the optimization with differential microphone arrays need to be re-normalized by

$$w_n = \hat{w}_n \frac{2}{2n+1}, \quad (2.41)$$

where the factor $\frac{2}{2n+1}$ is defined by the energy of the Legendre polynomials, see orthogonality property (2.14) or (2.43).

Consequently, the presented beamforming system is capable of generating steerable cardioid, supercardioid and hypercardioid shaped pick-up patterns of order N by setting the order weightings w_n according to table 2.3.

Type	Order	DI (dB)	F (dB)	Beamwidth	Nulls	w_n
Hyper-cardioid	1	6	8.5	105°	109°	(1)
	2	9.5	8.5	66°	73°, 134°	(1, 1)
Super-cardioid	1	5.7	11.4	115°	125°	(0.577)
	2	8.3	24	80°	104°, 144°	(0.775, 0.4)
Cardioid	1	4.8	8.5	131°	180°	(0.333)
	2	7	14.9	94°	180°	(0.5, 0.1)

Table 2.3: Weighting coefficients [Mor06] and quality measures [HB04] of pick-up patterns generated with spherical harmonics.

The weights for higher order beamforming can be found in [Mor06]. Note that the generated beam-patterns are not normalised to their maximal sensitivity, meaning $\sum_n w_n \neq 1$.

2.2.3 Discrete spherical arrays

In practice, the sound field is sampled at the microphone positions $\mathbf{r}_j = r\boldsymbol{\theta}_j$, where $j = 0, \dots, M-1$ denotes the j^{th} microphone on the sphere with radius r . Consequently the maximum order of spherical harmonics and thus the order of the designed beam patterns is limited to the maximum order N.

The array directivity for the sampled case can be calculated by

$$\begin{aligned} d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) &= \sum_{j=0}^{M-1} \alpha_j x(k\mathbf{r}, \boldsymbol{\theta}_j) g(k, \boldsymbol{\theta}_j) \\ &= \sum_{n=0}^N \sum_{m=-n}^n \chi_n^m(k\mathbf{r}, \boldsymbol{\theta}_i) g_n^m(k\mathbf{r}, \boldsymbol{\theta}_1), \end{aligned} \quad (2.42)$$

where $\chi_n^m(k\mathbf{r}, \boldsymbol{\theta}_j)$ and $g_n^m(k\mathbf{r})$ are the spherical harmonic coefficients of the sampled sound field and the sampled weighting function respectively and α_j is a real value which depends on the sampling scheme and the dependence of the direction $\boldsymbol{\theta}_i$ is omitted as it is not explicitly required. Note that eq. (2.42) is only valid if the discrete orthonormality condition is satisfied, see appendix A.4. For a uniform sampling scheme, where the microphones are equidistantly spaced the discrete orthonormality condition is satisfied when setting $\alpha_j = \frac{4\pi}{M}$. Note that true equidistant spatial sampling is only possible if the sampling points are arranged according to the Platonic solids [YSS⁺11]. Another possibility is to arrange the sampling points according to a spherical t-design [HS96]. This design method gives a distribution of sampling points such that the sampled spherical harmonics and thus the Legendre functions remain orthogonal on the sphere

$$\int_{S^2} P_n(\langle \boldsymbol{\theta}_i, \boldsymbol{\theta} \rangle) P_{n'}(\langle \boldsymbol{\theta}_i, \boldsymbol{\theta} \rangle) d\boldsymbol{\theta} = \frac{4\pi}{M} \sum_{j=0}^{M-1} P_n(\langle \boldsymbol{\theta}_i, \boldsymbol{\theta}_j \rangle) P_{n'}(\langle \boldsymbol{\theta}_i, \boldsymbol{\theta}_j \rangle) = \frac{4\pi}{2n+1} \delta_{nn'}. \quad (2.43)$$

But eq. (2.43) is only valid up to $2N \leq t$, so for example a 5-design allows for analysis with $N = 2$. However, the number of microphones M has to be larger or equal to $(N+1)^2$ to avoid spatial aliasing [RWB07].

With all the above relations and with eq. (2.42) the array output of a spherical harmonic beamforming system with look direction $\boldsymbol{\theta}_1$ is determined by

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_1) = \sum_{n=0}^N \sum_{m=-n}^n \chi_n^m(k\mathbf{r}, \boldsymbol{\theta}_i) \frac{1}{4\pi i^n \rho_n(k\mathbf{r})} Y_n^m(\boldsymbol{\theta}_1) w_n, \quad (2.44)$$

where w_n are the order weights that are designed to optimize the beampattern for specific constraints.

Ideal beams and DI: Figure 2.9 shows order-limited pick-up patterns that are possibly generated with a spherical microphone array beamforming system. These beams can be steered to any direction in the three-dimensional space, are ideally frequency independent and obey a constant directivity index (DI) over all steering directions $\boldsymbol{\theta}_1$, see table 2.3.

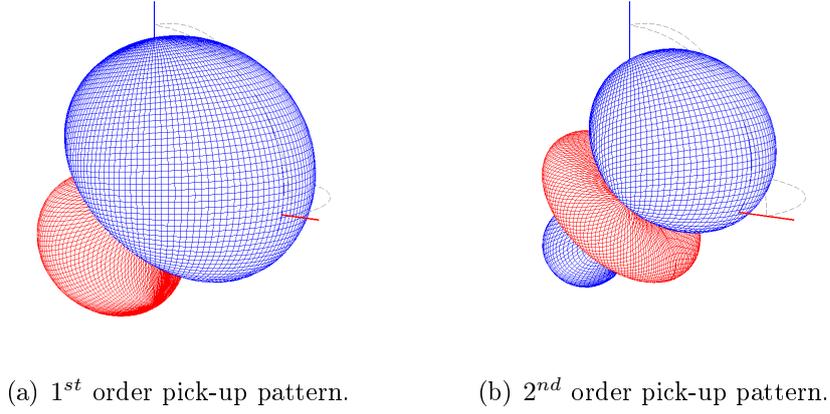


Figure 2.9: 3D hypercardioid pick-up patterns.

A formal definition of the DI is given in [BW01]

$$DI = 10 \log_{10} \left(\frac{4\pi |d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_1)|^2}{\int_{\boldsymbol{\theta}} |d(\boldsymbol{\theta}_1, \boldsymbol{\theta})|^2 d\boldsymbol{\theta}} \right) \quad (2.45)$$

and can be interpreted as the ratio of the gain for the lookdirection $\boldsymbol{\theta}_1$ to the average gain over all directions and thus it describes the ability to suppress the diffuse field.

With eq. (2.44) and the orthogonality of the spherical harmonics the DI can be rewritten to

$$DI = 10 \log_{10} \left(\frac{4\pi \left| \sum_{n=0}^N \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}_1) Y_n^m(\boldsymbol{\theta}_1) w_n \right|^2}{\left| \sum_{n=0}^N \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}_1)^2 w_n^2 \right|} \right). \quad (2.46)$$

2.2.4 Relation to hemispherical microphone arrays

As just the even-symmetric spherical harmonics (SH) with respect to the horizontal plane satisfy the boundary condition (velocity $\mathbf{v} = 0$ and pressure p is maximal) the skew symmetric harmonics are not considered, see fig. 2.10.

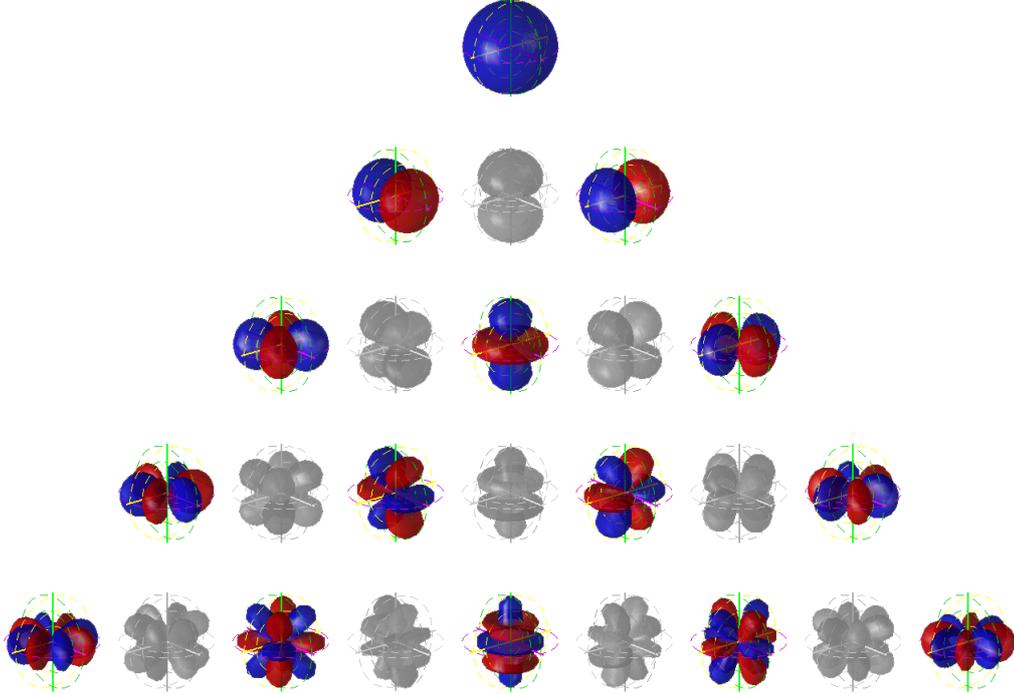


Figure 2.10: Spherical harmonics that satisfy the boundary condition (colored).

The transition between an analysis with the full set of spherical harmonics and a set where just the the even symmetric ones are used is expressed by changed summation indices. Equation (2.47) defines the summation indices for a summation over all orders and degrees that are available

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n c_n^m \implies \sum_{n=0}^{\infty} \sum_{l=0}^n c_n^{(2l-n)}, \quad (2.47)$$

where $(2l - n)$ is used instead of m to define the spherical harmonic degree. Thus, this means that the number of summands for a sum over a order n changes from $(2n + 1)$ to $(n + 1)$ if just the even symmetric spherical harmonics are used. The upper bound of summation for band limited functions is set to N (maximum order). In this case the total number of summands changes form $(N + 1)^2$ to $\frac{(N+1)(N+2)}{2}$.

Equation (2.48) defines the summation of spherical harmonic signals with the same

	Full set	Even symmetric
Sum over all orders and degrees	$(N + 1)^2$	$\frac{(N+1)(N+2)}{2}$
Sum over one order n	$2n+1$	$n+1$
Sum over one degree m	$N - m $	$\lfloor \frac{N- m }{2} \rfloor + 1$

Table 2.4: Number of summands (band limited functions).

degree m

$$\sum_{n=|m|}^N c_n^m \implies \sum_{l=0}^{\lfloor \frac{N-|m|}{2} \rfloor} c_{2l+|m|}^m. \quad (2.48)$$

Note that the variable c is a substitute variable for any variable that is related to the order n or degree m , as for example the spherical harmonics, the radial functions or the spherical wave spectrum.

A comparison of the number of summands is given in 2.4.

Ideal beams and DI: By substituting eq. (2.47) into eq. (2.44) the array output (pick-up pattern) for a scenario where just the even-symmetric real-valued SH are considered is given as

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_l) = \sum_{n=0}^N \sum_{l=0}^n \chi_n^{(2l-n)}(kr, \boldsymbol{\theta}_i) \frac{1}{4\pi i^n \rho_n(kr)} Y_n^{(2l-n)}(\boldsymbol{\theta}_l) w_n. \quad (2.49)$$

One can easily see from eq. (2.49) and fig. 2.10 that the shape of pick-up patterns varies as the zenith angle of the lookdirection ϑ_l changes.

Figure 2.11 shows the three dimensional beampatterns steered from the horizontal ($0^\circ, 90^\circ$) to vertical ($0^\circ, 0^\circ$) array plane. It can be seen that the shape of the pick-up patterns is not invariant with respect to the zenith of the look direction and therefore it is also evident that the directivity index changes as the zenith changes.

With eq. (2.47) and eq. (2.46) the DI for pick-up patterns generated with even-symmetric SH is calculated with

$$DI = 10 \log_{10} \left(\frac{4\pi \left| \sum_{n=0}^N \sum_{l=0}^n Y_n^{(2l-n)}(\boldsymbol{\theta}_l) Y_n^{(2l-n)}(\boldsymbol{\theta}_l) w_n \right|^2}{\left| \sum_{n=0}^N \sum_{l=0}^n Y_n^{(2l-n)}(\boldsymbol{\theta}_l)^2 w_n^2 \right|} \right). \quad (2.50)$$

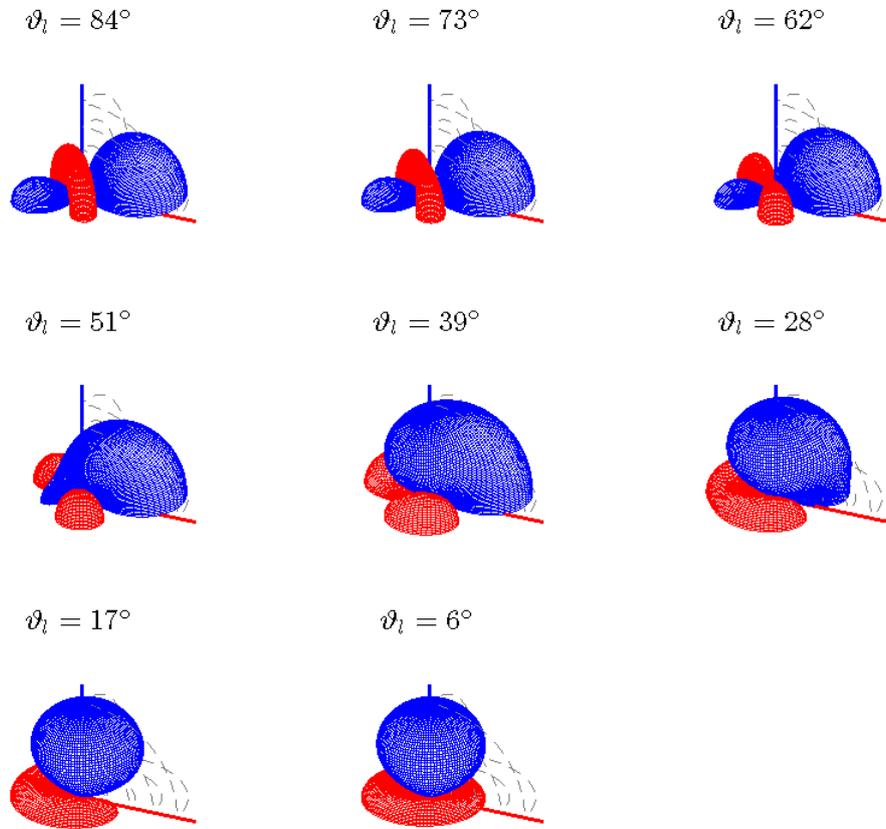


Figure 2.11: 3D 2nd order beam patterns (hypercardioid weighting).

The achievable DIs for different weightings and look directions is depicted in fig. 2.12 where it can be recognized that the hypercardioid weighting has the highest directivity. Further one can easily see the additional gained degree of freedom of second order beamforming when compared to first order beamforming, where it is not possible to steer a beam towards the z-direction.

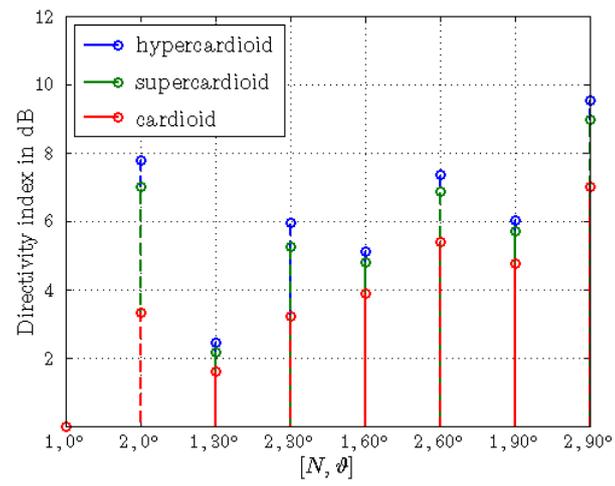


Figure 2.12: Directivity index of produced beams.

3. Measuring the characteristics of the prototype



Figure 3.1: Prototype of a second-order circular microphone array.

Figure 3.1 shows a prototype circular microphone array that is based on a theoretical concept established by Franz Zotter and Hannes Pomberger (IEM). Richard Pribyl and Marco Riemann (AKG) supervised the actual implementation of the prototype.

The array is planar and intended for use as an interfacial microphone that is placed for example on table. Possible applications include beamforming for teleconferencing and to a certain extent also 3D spatial recordings. The array consists of a central omnidirectional microphone and five cardioid microphones that are arranged equally spaced on a concentric circle with radius $r = 2.1\text{cm}$.

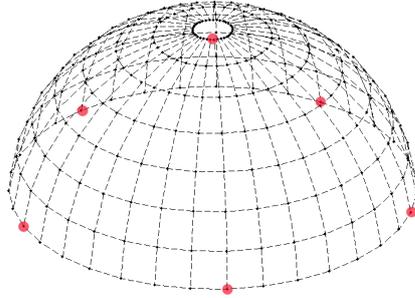
This section explains the necessary steps to measure and display the frequency dependent three-dimensional directivity patterns of the individual array microphones in their final assembly. As an initial step, we analyze the characteristics of the array and identify the directivity patterns of the mounted array sensors. In order to do so, we measure the impulse responses (IRs) of each microphone due to excitations originating from various directions. The simplest setup to do so would only consider loudspeakers at the six directions that are on axis to one of the microphones, each, see red dots fig. 3.2(a) and sec. 4.3.1. However, in an initial step a very dense set of excitation points is chosen since it allows for higher order resolution than two (see eq. (3.12)). The higher order resolution is essential, as the directivity patterns of the microphones potentially contain orders higher than one in the higher frequency region (above approx. 2kHz).

3.1 Measurement setup

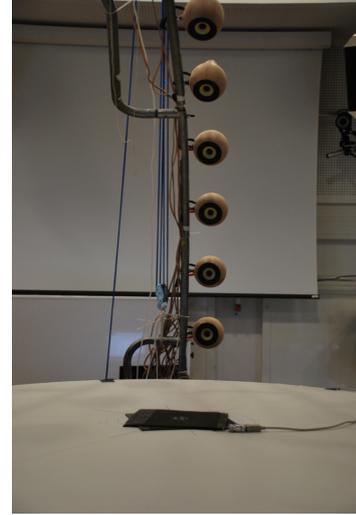
The transmission path between the six array-microphones and one specific excitation-direction can be measured by recording a measurement signal which is radiated from a loudspeaker that is placed in this direction. For a complete measurement of as many directional responses as possible, the direction dependent response of each microphone is measured by a hemispherical, concentric loudspeaker array surrounding.

Measurement excitation positions (loudspeaker positions) are set according to a spatial resolution of $\Delta\varphi = 10^\circ$ in azimuth and $\Delta\vartheta = 11.25^\circ$ in zenith direction (see spherical coordinates (A.1)) which leads to a grid layout of 8 latitude circles and 36 meridians. Figure 3.2(a) depicts the chosen measurement configuration of 288 transducers distributed on a half sphere surface with a radius of $1.3m$ centred around the circular array. In order to reduce the measurement effort the number of excitation points may be reduced to 6 instead of 288, see red dots in fig. 3.2(a).

As shown in fig. 3.2(b), an eight-element quarter-circular loudspeaker array could be used for sequentially measuring 8 directions of different zenith angles at one azimuth angle. In order to reach all 36 proposed azimuth angles the microphone array was rotated by a computer controlled turntable, after all measurements for one azimuth were done.



(a) Schematic measurement setup consisting of 8 latitude circles with $\Delta\theta = 11.25^\circ$ and 36 equidistant meridians spaced with $\Delta\varphi = 10^\circ$.



(b) Measurement setup

Figure 3.2: Measurement setup.

The measurements span a $i = 1, \dots, 8 \times 36$, $j = 1, \dots, 6$ and $n = 0, \dots, N - 1$ set of impulse responses $h_{ij}[n]$, where i , j and n are the loudspeaker, microphone and discrete-time indices, respectively. Note that the symbols n , and N as they are used here must not be confused with the order index and maximum order used in other sections.

Method: The actual measurements were done according to the exponentially swept sine (ESS) method presented by Farina in [Far00]. This technique was chosen because it provides advantages over other well known measurement techniques such as the linear sine sweep method, the maximum length sequence (MLS) method, as well as other techniques that measure the IR directly [SEA02]. The ESS method uses an exponentially-swept sine for excitation and an aperiodic deconvolution to extract the impulse response from the recorded sweep response of the transmission path and is explained shortly throughout the next paragraph.

Suppose that $\hat{s}_{ij}[n]$ denotes the recorded response of the transmission path between the i^{th} loudspeaker to the j^{th} microphone to an exponential reference sweep denoted by $s[n]$. Then the impulse responses are calculated in the frequency domain by a

simple division of the two spectra defined in eq. (3.1)

$$h_{ij}[n] = IFFT \left[\frac{FFT(\hat{s}_{ij}[n])}{FFT(s[n])} \right]. \quad (3.1)$$

In order to suppress artefacts of cyclic deconvolution, a DFT length was chosen that is large enough (for example $2N$, where N corresponds to the longer of the two signals.)

3.1.1 Compensation of the transducer frequency response

Figure 3.3 shows the whole measurement chain. Accordingly, the Fourier Transform of the measured signal is compromised by an interaction of the frequency responses of the actual signal, the D/A and A/D converters, the pre-amps, the loudspeaker, the room and the microphone. If the conversion and the pre amplification are assumed to be ideal then they may be neglected. Then the Fourier Transform of the measured signal can be defined by eq. (3.2).



Figure 3.3: Measurement chain

$$Y(\omega) = X(\omega) \cdot H_{Ls}(\omega) \cdot H_{Room}(\omega) \cdot H_{Mic}(\omega) \quad (3.2)$$

The influences of the loudspeaker characteristics are minimized by equalizing according to a reference measurement (the placement of the reference microphone can be seen in fig.: 3.4).

Figure 3.5 shows the designed equalization filter, where the magnitude response of the equalization filters is limited to a dynamic range of 20dB and the according phase response is realized as a minimum-phase system.

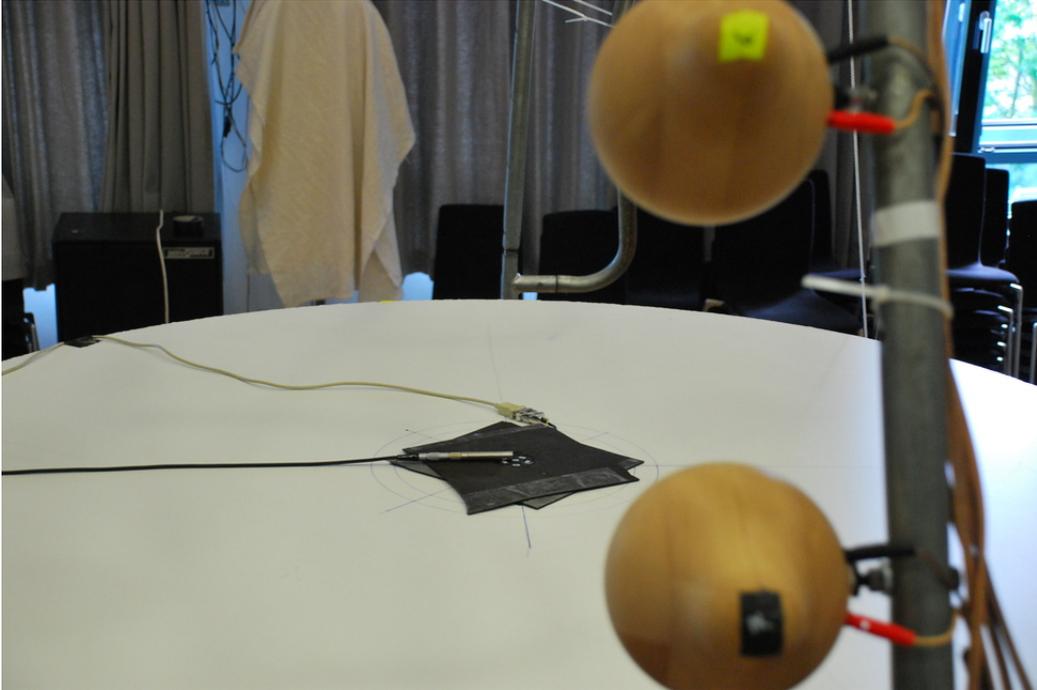


Figure 3.4: Placement of reference microphone.

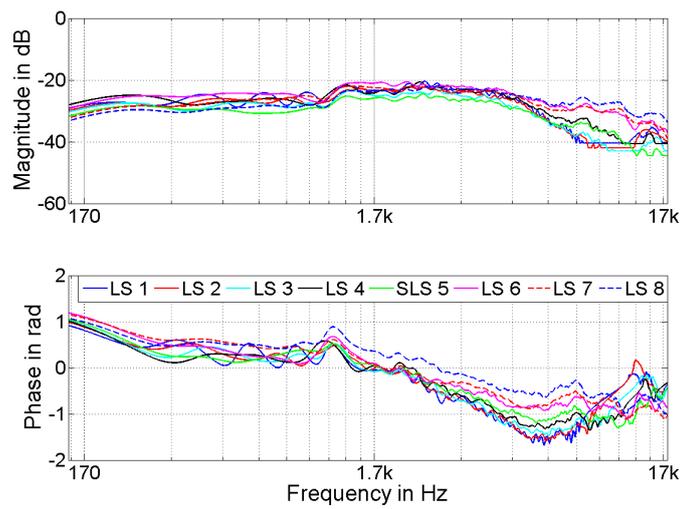


Figure 3.5: Equalization filters.

3.1.2 Measured impulse responses

Figure 3.6 shows the IRs of all six microphones for one specific excitation position defined by $(\varphi, \vartheta) = (0^\circ, 50.625^\circ)$. The measurements are performed in the IEM CUBE [ZRS03] which violates the demand of a by default anechoic measurement room. Therefore, the obtained IRs have to be windowed in order to suppress the parts of the response which are caused by the reflections of the room and other objects in the measurement setup. A window of 256 samples is used (@ $f_s=44100\text{Hz}$ this equals a time of 6 ms; so reflections with a path difference of 2m compared to the direct path are suppressed; in order to obtain a valid result, at least one period of the lowest frequency has to fit inside the time window of 6 ms, which leads to a bottom frequency of $f_u = 170\text{ Hz}$). Figure 3.7 depicts the magnitude and the phase response of the IRs. Due to the microphone spacing the responses show pairwise evolutions (cf. fig. 3.1).

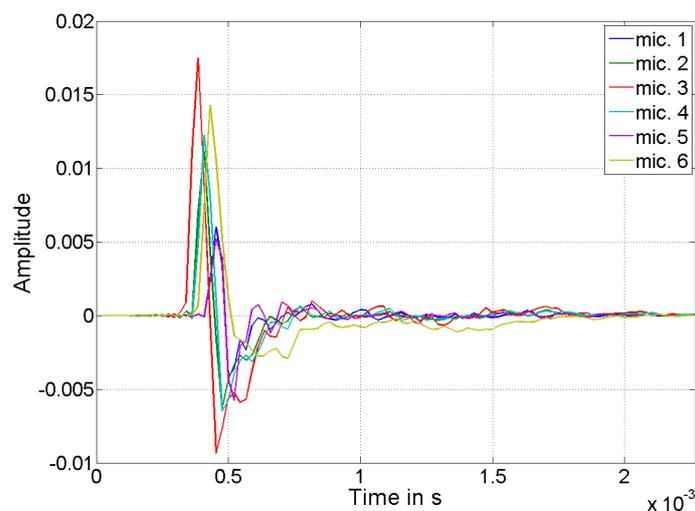


Figure 3.6: Impulse responses of all 6 microphones for a source located at $\varphi = 0^\circ$ and $\vartheta = 50.625^\circ$

3.1.3 Time realignment

A small displacement of the centred array or the surrounding loudspeaker circle could cause slightly changed sphere radii. In order to minimize the thereby introduced phase

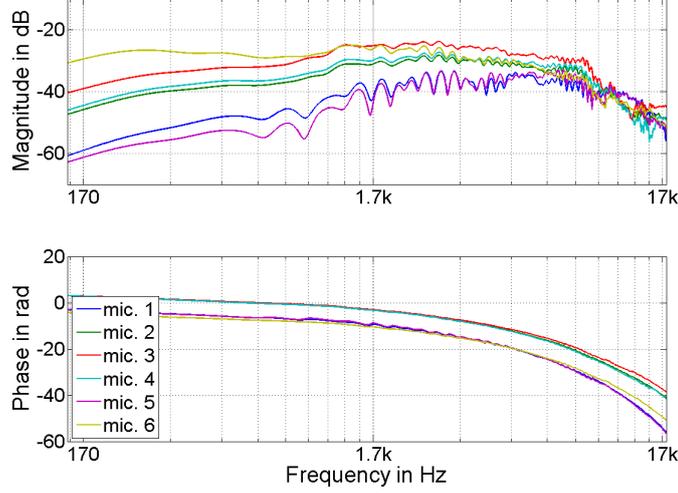


Figure 3.7: Frequency responses of all 6 microphones. The upper plot shows the magnitude and the lower shows the unwrapped phase for a source located at $\varphi = 0^\circ$ and $\vartheta = 50.625^\circ$.

errors, the impulse responses are realigned in time. Note that the amplitude errors are considered to be negligible since the distance to the loudspeaker is many times greater than the maximally assumed displacement (cf. 1.3m vs. ± 0.05 m).

The centred omnidirectional microphone is used as a reference for time re-alignment since it marks the acoustical centre point of the array. Precisely, the maximum value in the impulse response was used to detect the delay differences that are necessary for the realigning procedure. In order to increase alignment accuracy the signals are up-sampled by a factor of 4.

The actual shifting is applied to the up-sampled signal by multiplying the spectra with a linear phase term. Equation (3.3) illustrates the procedure

$$\widehat{H}_{ij}(k) = H_{ij}(k) e^{-i(2\pi \frac{k}{N} \cdot \text{delay}(i))}, \quad (3.3)$$

where k defines the frequency bin, N the DFT size and $\text{delay}(i)$ the desired delay in samples that is used for the shifting of all microphone signals $j = 1, \dots, 6$ for one loudspeaker position θ_i .

For verification figure 3.8 shows the up-sampled IRs of the central microphone for the original and realigned case. It can be seen that the primary differences are reduced

by time-realignment.

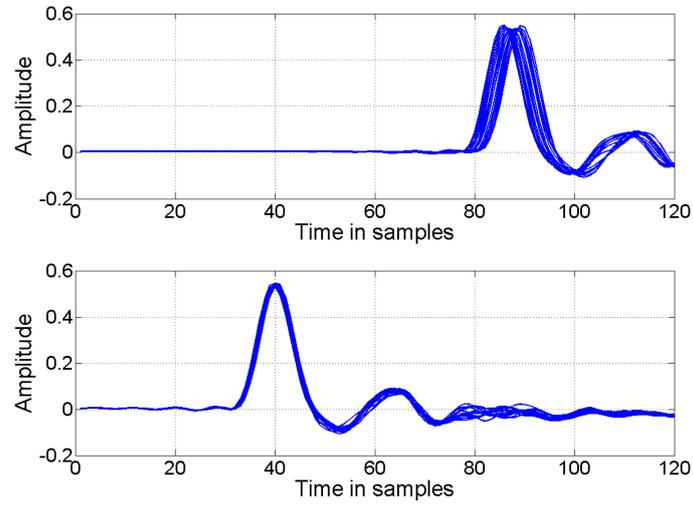


Figure 3.8: (top) original and (bottom) time realigned IRs (up-sampled by 4).

3.2 Directivity patterns

In this section we analyse the measurement data in order to make a statement about the three dimensional directivity properties and the rotational symmetry of the analysed microphones characteristics.

3.2.1 Displaying at discrete directions

Let us define a vector that contains the discrete directional response of the j^{th} microphone as

$$\mathbf{h}_j(\omega) = \begin{pmatrix} h_j(\boldsymbol{\theta}_1, \omega) \\ h_j(\boldsymbol{\theta}_2, \omega) \\ \vdots \\ h_j(\boldsymbol{\theta}_L, \omega) \end{pmatrix}, \quad (3.4)$$

where θ_i denotes the direction of the i^{th} loudspeaker with $i = 1, \dots, L$. The magnitude of the directivity pattern at a specific frequency is then plotted at the available grid positions (here linearly connected) in fig. 3.9.

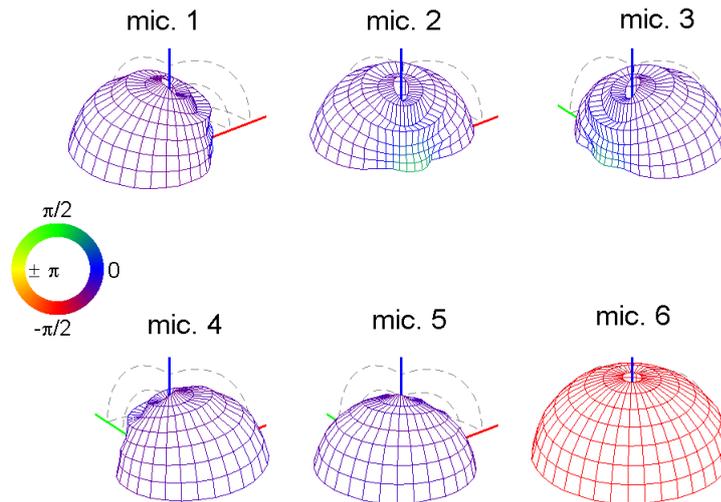


Figure 3.9: Directivity patterns of the microphones at a frequency of $f = 1077$ Hz.

Obviously, the microphones 1 to 5 sit on the concentric circle and their cardioid

directivity is facing outwards. Microphone 6 is in the centre and omnidirectional. Due to the limited number of measurement directions, the plots look rather coarse. For a closer investigation spherical-harmonic interpolation is applied.

3.2.2 Interpolated directivity patterns

A compact overview of the needed mathematical and physical fundamentals is given in the appendix A.

Throughout this section about the spherical-harmonic interpolation all reference to the frequency dependence is omitted because it is not explicitly required.

Starting with the vector \mathbf{h} that contains the measured IRs it is possible to select a single impulse response by introducing a vector $\mathbf{g} = [g_1, \dots, g_L]^T$ that corresponds to the driving voltages of the loudspeakers

$$h_j(\boldsymbol{\theta}_i) = \mathbf{h}_j^T \mathbf{g}, \quad (3.5)$$

where $\mathbf{g} = [\delta_{ii}]$ which means that just the impulse response from the i^{th} loudspeaker to the j^{th} microphone is selected.

The determination of the driving voltages that allow for determining the impulse response from any direction is explained in the following paragraph.

Let us define a driving distribution $g(\boldsymbol{\theta})$ that is described by a sum of Dirac delta functions located at the centre of each loudspeaker weighted by the corresponding excitation signal g_i as

$$g(\boldsymbol{\theta}) = \sum_{i=1}^L \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i) g_i. \quad (3.6)$$

According to eq. (A.18) and [PZ08] the modal source strength coefficients are determined with

$$\phi_n^m = \sum_{i=1}^L Y_n^m(\boldsymbol{\theta}_i) g_i. \quad (3.7)$$

Assuming the system is band-limited on the sphere with maximum order N it is feasible to rewrite the band limited excitation control system in a matrix notation

$$\boldsymbol{\phi}_N = \mathbf{D}_N \mathbf{g}, \quad (3.8)$$

where the loudspeaker spherical-harmonics matrix is defined as

$$\mathbf{D}_N = \begin{bmatrix} \mathbf{y}_N^T(\boldsymbol{\theta}_1) \\ \mathbf{y}_N^T(\boldsymbol{\theta}_2) \\ \vdots \\ \mathbf{y}_N^T(\boldsymbol{\theta}_L) \end{bmatrix}. \quad (3.9)$$

The driving voltages are finally determined by applying the pseudo inverse

$$\mathbf{g} = \mathbf{D}_N^+ \boldsymbol{\phi}_N, \quad (3.10)$$

where \mathbf{D}_N^+ provides the mapping of the modal source strength excitation vector onto the driving voltages and is referred to as decoder matrix. By setting $\boldsymbol{\phi}_N = \mathbf{y}_N(\boldsymbol{\theta})$ one obtains the driving voltages that allow for accessing the spherical-harmonic interpolated impulse response corresponding to an excitation from direction $\boldsymbol{\theta}$

$$h_j(\boldsymbol{\theta}) = \mathbf{h}_j^T \mathbf{D}_N^+ \mathbf{y}_N(\boldsymbol{\theta}). \quad (3.11)$$

Please note that due to the boundary condition only the evensymmetric spherical harmonics with respect to the horizontal plane are considered. In this case the relation between the sampling points L and the maximal spherical harmonic order N is given as

$$L \geq \frac{(N+1)(N+2)}{2}. \quad (3.12)$$

With $L = 288$ sampling points and eq. (3.12) the number of controllable spherical harmonics for an equidistant grid layout would be limited to $N = 22$. As the the grid layout under consideration approximately has equal angular distributed sampling points it is according to [PZ08] possible to use the conditioning of $\kappa(\mathbf{D}_N^T \mathbf{W} \mathbf{D}_N)$ as an indicator for setting the maximal order N . Thereby, \mathbf{W} is a diagonal weighting matrix containing whether weightings that are equivalent to the surface fraction of the corresponding excitation point or equal weights (cf. t-design). One can easily observe from fig 3.10 that the conditioning for both schemes skyrockets for orders $N > 15$ and thus the maximal order is set to $N = 14$ in practice. Further the conditioning of the surface fraction weighting is preferable.

Summarized the spherical-harmonic interpolation allows to evaluate the directional characteristics at a smoother angular grid. For an equiangular interpolated grid with

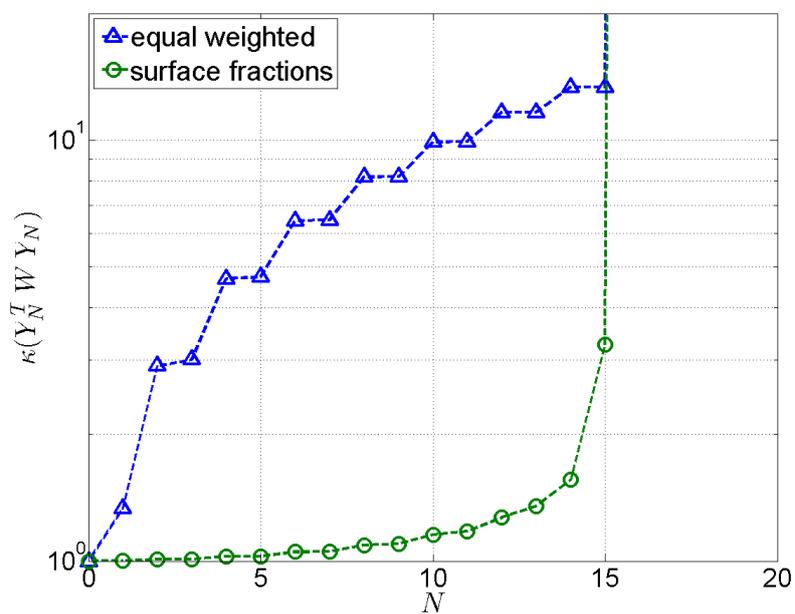


Figure 3.10: Conditioning of the pseudo inverse using equal weights or surface fraction dependent weights for the discussed grid layout 3.2(a).

$\Delta\varphi = 2.4^\circ$ and $\Delta\vartheta = 2.25^\circ$, i.e. 12000 nodes can be inserted. The figures 3.11 and 3.12 exemplarily show the three dimensional directivity of the microphones 3 and 6 for one specific frequency. The phase information of the transformed IRs is described by the colour gradient and can be interpreted if one regards the coloured phase circle in the figure.

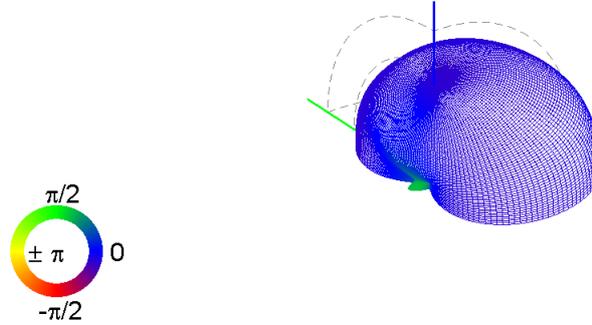


Figure 3.11: Three dimensional directivity pattern of microphone 3 at $1077Hz$.

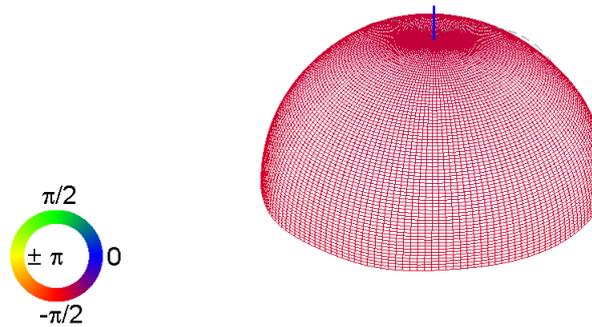


Figure 3.12: Three dimensional directivity pattern of microphone 3 for $1077Hz$.

3.2.3 Slices of the three dimensional directivity patterns

In order to analyze the rotational symmetry and for sake of observation simplicity, the three dimensional directivity patterns are cut in slices, see fig. 3.13. In an initial step the patterns of the cardioid microphones are rotated around the z-axis to guarantee the highest sensitivity of each microphone for an azimuth angle of $\varphi = 0^\circ$. Next the cutting plane is rotated around the z-axis by the rotation angles $\alpha_r = (0^\circ + r \Delta\alpha)$ with $r = 0 : 6$ and $\Delta\alpha = 30^\circ$; cf. figure 3.13(a).

Figure 3.14 depicts slices for four rotation angles overlaid in one plot. The directivity patterns of tow exemplary cardioid microphones are strongly frequency dependent (cf. figures 3.14(a) and 3.14(c)). Whilst the microphones on the circle exhibit approxi-

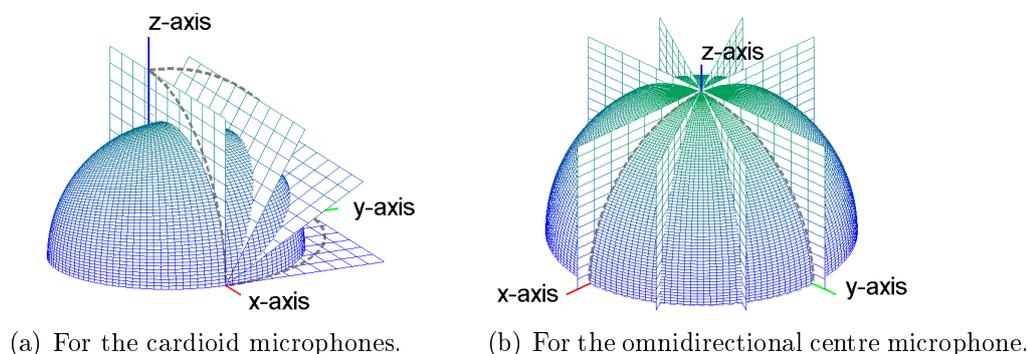


Figure 3.13: Scheme of the cutting plane orientations.

mately cardioid characteristics in the lower octave band with $f_c = 300$ Hz, they deform their directional sensitivity for higher frequencies in the band around $f_c = 2000$ Hz. Further, a strong sensitivity decay towards sound impinging from $\vartheta = 0^\circ$ attracts attention, see fig. 3.14(b).

In order to analyze the characteristics of the omnidirectional centre microphone the cutting planes are defined according to figure 3.13(b). The slices for the octave bands with centre frequencies f_c are illustrated in figure 3.15 for the case of vertical slices. The measured directivity patterns share the same phase at all angular positions. Further the directivity is symmetric. As expected the phase evolution or the phase differences for the omnidirectional microphone is not changing over the different angles.

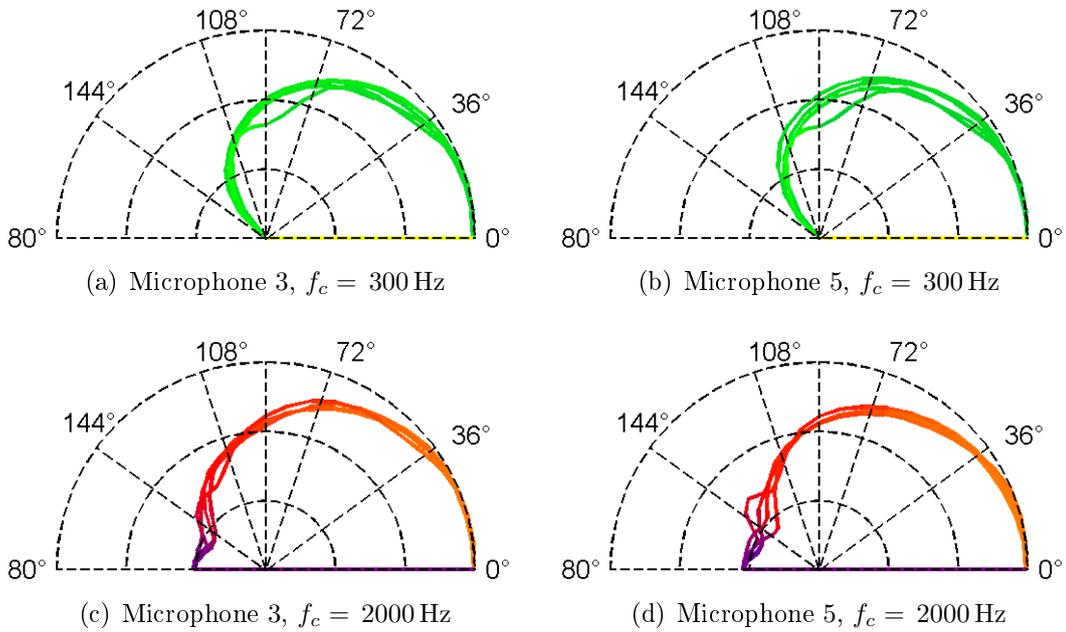


Figure 3.14: Slices evaluated for an octave band with center frequency f_c at rotation angles $\alpha = (0^\circ, 30^\circ, 60^\circ, 90^\circ)$. Dynamic range 30dB with 10dB/division.

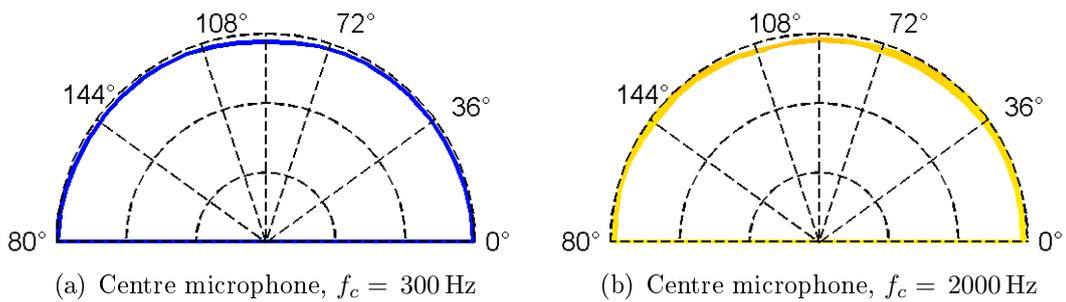


Figure 3.15: Slices through the omnidirectional centre microphone and octave bands with centre frequency f_c . Dynamic range 30dB with 10dB/division.

4. Modal SH beamforming based on measurements

This chapter deals with the implementation of a modal SH beamforming system for the described array prototype presented in chapter 3.

The general structure of a modal spherical harmonics beamformer is shown in fig. 4.1. It is characterized by an elegant separation of a decomposition or more general an analysis unit and a beamforming unit [ME02]. Generally, the decomposer is respon-

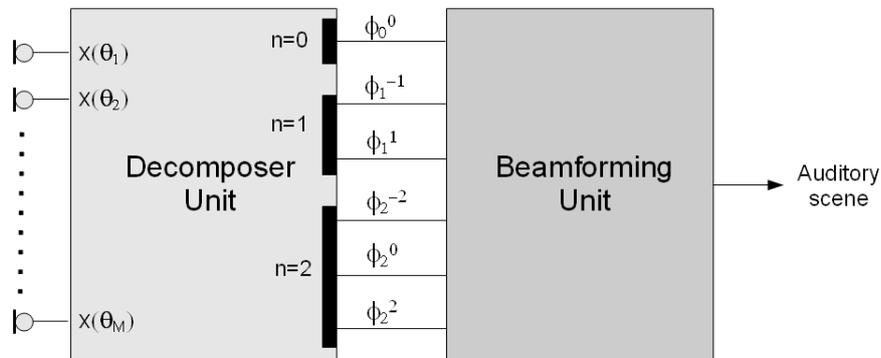


Figure 4.1: Scheme of a modal beamforming approach.

sible for transforming the sound field, which is sampled by the microphones, into the spherical harmonics domain. As the aim is to design frequency independent beam-patterns the spherical harmonic coefficients $\chi_n^m(kr)$ of the microphone signals have to be normalized using the inverse of the radial and frequency dependent term $\rho_n(kr)$ of the solution, see eq. (2.44). The obtained coefficients of the incident wave-field b_n^m are next used to evaluate the modal source amplitudes, see eq. (A.32). The actual

beamforming is then realized by weighting and summing the modes ϕ_n^m [ZPF09]. The modal source strength is used for the more general case of a surround distribution of plane waves instead of the considered case of a single plane wave (see ch. 2) and thus, $Y_n^m(\theta_i)$ is substituted by ϕ_n^m in eq. (2.40).

However, an approach where the decomposition of the sound-field is not based on an analytic model, but on measured data is presented throughout this chapter.

4.1 Decomposition based on measurements

As described in chapter 3 it is possible to determine the pick-up patterns of all microphone array sensors using an array of loudspeakers on a concentric hemisphere surrounding the microphone array, i.e transfer function between loudspeakers and microphones, see fig. 3.2. How to design a system that transforms the recorded microphone signals into the spherical harmonic spectrum ϕ_n^m , which are the target signals for spherical harmonics based beamforming, is explained in the following section.

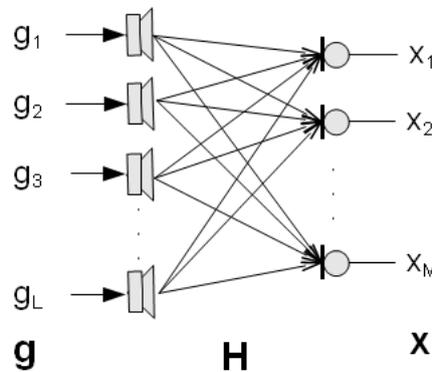


Figure 4.2: MIMO system.

The multiple-input-multiple-output system (MIMO, see fig. 4.2) of the device under test is described as

$$\mathbf{x}(\omega) = \mathbf{H}(\omega) \mathbf{g}(\omega), \quad (4.1)$$

where $\mathbf{H}(\omega)$, $\mathbf{x}(\omega)$ and $\mathbf{g}(\omega)$ represent the impulse responses, the microphone signals and the loudspeaker signals in the frequency domain, respectively. The matrix \mathbf{H} is

defined as

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_M^T \end{pmatrix}, \quad (4.2)$$

where \mathbf{h}_j is defined in eq. (3.4). Note that the dependency on the frequency is omitted for better readability.

In order to design a system that allows for a correct spherical harmonic decomposition of the sound field it is necessary to invert the MIMO system, where the angularly distributed loudspeaker driving voltages \mathbf{g} are represented in the spherical harmonics domain. The transformation of the right side of MIMO system is already defined in section 3, where the matrix \mathbf{D}^+ is used for spherical-harmonic interpolation. By substituting eq. (3.10) into eq. (4.1), the MIMO system with a transformed input side is defined as

$$\mathbf{x} = \mathbf{H} \mathbf{D}_{N_C}^+ \phi_{N_C}. \quad (4.3)$$

Direct inversion: Presumably the most obvious solution in order to obtain a exact decomposition is to invert the system defined in eq. (4.3)

$$\hat{\phi}_{N_A} = (\mathbf{H} \mathbf{D}_{N_C}^+)_{N_A}^+ \mathbf{x} = (\mathbf{H} \mathbf{D}_{N_C}^+)_{N_A}^+ \mathbf{H} \mathbf{D}_{N_C}^+ \phi_{N_C}. \quad (4.4)$$

Note that the maximal decomposition order $N_A = 2$ needs to be introduced as a microphone array consisting of 6 sensors is not applicable to resolve a sound field with higher spatial frequency as N_A , see chapter 5. From eq. (4.4) one can easily see that the multiplication of the microphone signals with the inverted system allows for a correct decomposition of the sound field up to order N_A if $(\mathbf{H} \mathbf{D}_{N_C}^+)_{N_A}^+ \mathbf{H} \mathbf{D}_{N_C}^+ = [\mathbf{I}, \mathbf{0}]$. Note that a direct inversion does not require any model on the microphone side.

Inversion of the transformed MIMO system: The transformation of the microphone signals in the spherical harmonics domain is done by applying the discrete spherical harmonics transform (DSHT) as explained in the appendix A.5. Let us define the microphone signals using spherical harmonics and matrix notation as

$$\mathbf{x} = \mathbf{C}_{N_A} \chi_{N_A}, \quad (4.5)$$

where \mathbf{C}_{N_A} denotes the spherical harmonics matrix evaluated at the microphone positions $\boldsymbol{\theta}_j$ and $\boldsymbol{\chi}_{N_A}$ denotes the corresponding spherical wave spectra. By substituting eq. (4.5) into eq. (4.3) the both-sided transformed MIMO system is defined as

$$\boldsymbol{\chi}_{N_A} = \underbrace{\mathbf{C}_{N_A}^+ \mathbf{H} \mathbf{D}_{N_C}^+}_{\dot{\mathbf{H}}_{N_A, N_C}} \boldsymbol{\phi}_{N_C}, \quad (4.6)$$

where $\dot{\mathbf{H}}_{N_A, N_C}$ comprises the complete MIMO system from the excitation to the measurement position in the spherical harmonics domain and therefore, it is referred to as system matrix. The subscript N_A, N_C is used as an indicator of the matrix size and throughout this section dotted variables indicate the spherical harmonics domain.

By adding the frequency independent encoder matrix $\mathbf{C}_{N_A}^+$ we expect a better separation of the spherical harmonics at the microphone side, c.f. relation between the wave spectrum, the spherical wave spectrum $\chi_n^m = b_n^m j_n(kr)$ and the spherical harmonics spectrum $b_n^m = -\phi_n^m i k h_n^{(2)}(kr_i)$. And thus, it is assumed that the inverted system contains non-relevant entries that may be omitted and consequently, the number of necessary decomposition filters is assumed to decrease when compared to the direct inversion.

Ideally, the system matrix $\dot{\mathbf{H}}_{N_A, N_A}$ should be a diagonal matrix (in the array achievable subspace), but one can easily see from fig 4.3 that this is not the case.

The frequency dependent evolution of the system matrix is shown in fig. 4.4, where the subscripts in the legend denote the row and column. Again one can easily see that the system matrix is not a diagonal matrix over the entire frequency range as ideally expected. Thus, the desired separation of the different excited modes is not fully achieved by inserting the encoder matrix $\mathbf{C}_{N_A}^+$.

However, the sound-field is correctly decomposed by inserting a decomposition matrix $\dot{\mathbf{R}}_{N_A, N_A}$ that is ideally left-inverse to the system in the spherical harmonics domain

$$\hat{\boldsymbol{\phi}}_{N_A} = \dot{\mathbf{R}}_{N_A, N_A} \dot{\mathbf{H}}_{N_A, N_C} \boldsymbol{\phi}_{N_C}. \quad (4.7)$$

According to [PZ08] the inversion of the system matrix can be done exact in the array achievable subspace where the system is truncated by N_A leading to a $(N_A + 1)(N_A +$

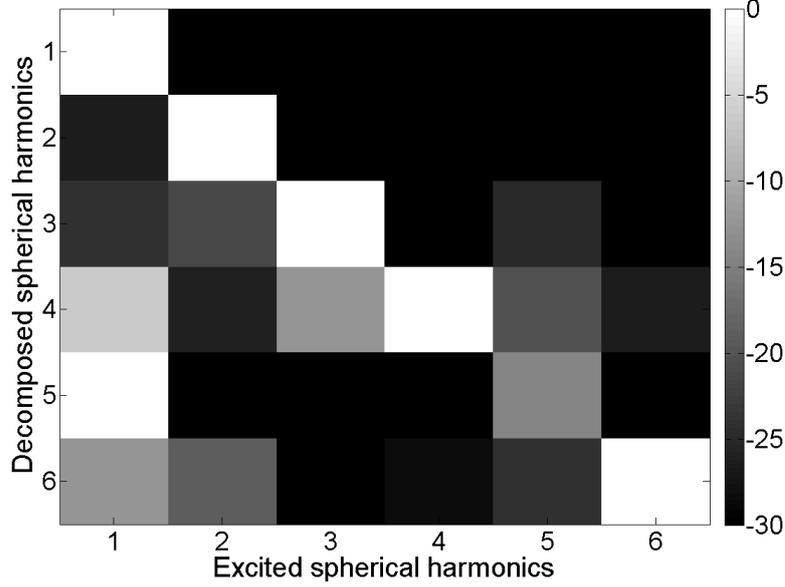


Figure 4.3: System matrix $\dot{\mathbf{H}}_{2,2}$ (6×6 subspace) for a frequency of $f = 2000 \text{ Hz}$ in dB . The values are limited to bandwidth of 30 dB and are normalized for each line separately.

$2)/2 \times (N_A + 1)(N_A + 2)/2$ square system matrix and a decomposition matrix defined to

$$\dot{\mathbf{R}}_{N_A, N_A} = (\mathbf{C}_{N_A}^+ \mathbf{H} \mathbf{D}_{N_A}^+)^{-1} \quad (4.8)$$

or by an approximate inversion of the system considering also the higher order harmonics at the excitation side

$$\dot{\mathbf{R}}_{N_A, N_A} = (\mathbf{C}_{N_A}^+ \mathbf{H} \mathbf{D}_{N_C}^+)^+_{N_A}. \quad (4.9)$$

In the second case it is also feasible to use only the $(N_A + 1)(N_A + 2)/2$ upper rows so that the control matrix exhibits the same size for the different calculation possibilities.

Under the assumption that the present wave field is strictly band limited with the spatial frequency N_A it is possible to perform a perfect decomposition by emphasizing eq. (4.8). On the other hand the inversion in the truncated subspace does not consider components of higher order harmonics for $N > N_A$ and thus occurring spatial aliasing corrupts the decomposed modes (cf. spatial aliasing A.5.2). The second approach (4.9) allows for an attenuation of harmonics that could cause spatial aliasing but with the drawback, that the calculation requires the approximate inversion of an over-

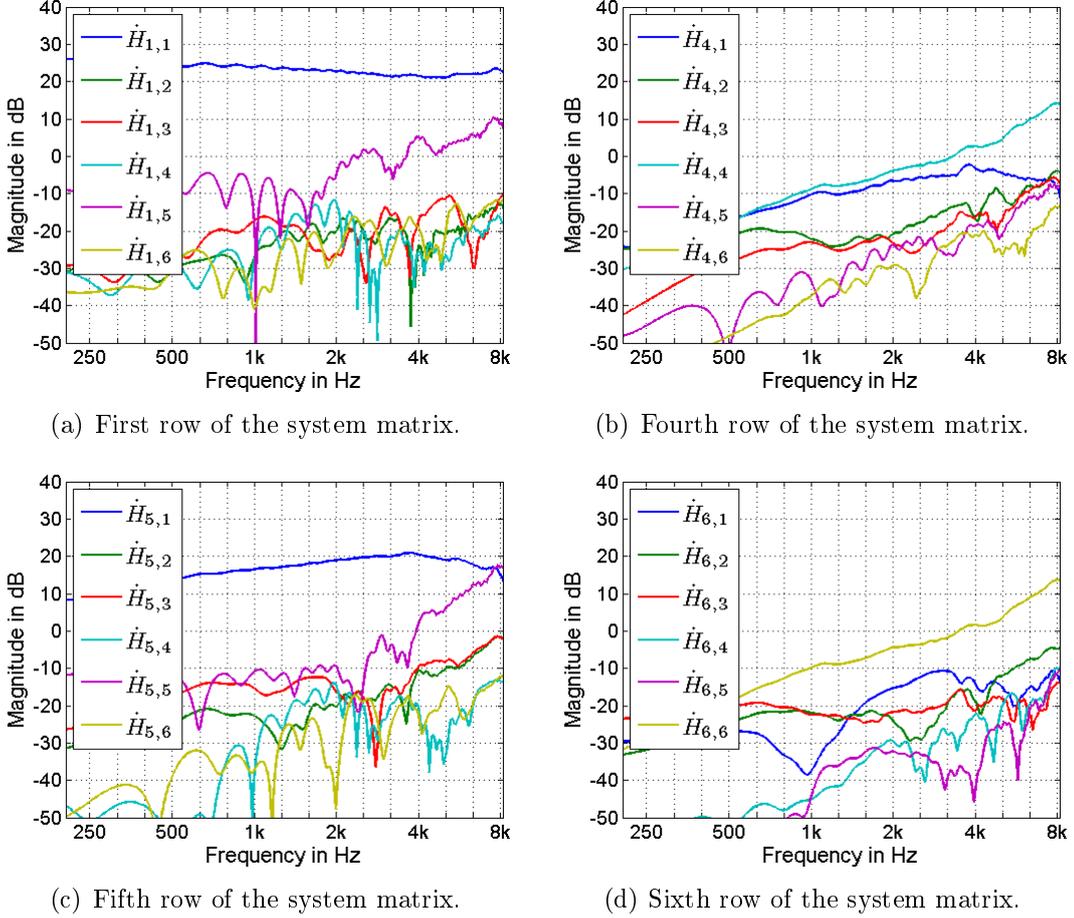


Figure 4.4: System matrix between 50 – 8000 Hz.

determined system and accordingly even strictly band limited wave-fields can not be decomposed exactly. However, we decided to generate the decomposition matrix according to (4.8) in order to guarantee for an exact reconstruction of the excited spherical harmonics in the achievable subspace.

By substituting (4.6) in (4.7) the decomposition process is written to:

$$\hat{\phi}_{N_A} = \mathbf{R}_{N_A, N_A} \mathbf{C}_{N_A}^+ \underbrace{\mathbf{H} \mathbf{D}_{N_A}^+}_{\mathbf{x}} \phi_{N_C} \quad (4.10)$$

In words this means that the recorded microphone signals x are initially multiplied with a frequency independent encoder $\mathbf{C}_{N_A}^+$ and that the aliased and frequency dependent modes χ_n^m are separated and unified using the frequency dependent decom-

position matrix $\dot{\mathbf{R}}_{N_A, N_A}$ to yield the correct or approximated modal source amplitude spectrum $\hat{\phi}_{N_A}$.

4.1.1 Practical inversion

The following paragraphs examine the invertibility and the accuracy of the system matrix inversion (eq. (4.8)) and give general background on the issue.

In general an arbitrary square matrix \mathbf{A} can be exactly inverted if it is non-singular or synonymical regular. As the condition number can be interpreted as an indicator for the distance of matrix to a regular matrix, the condition number of the system matrix is examined to make a statement about the inversion accuracy thereof. A preferable condition number of a system would be around $\kappa(\mathbf{A}) = 1$ which indicates a well-conditioned ¹ matrix whereas large values indicate an ill-conditioned matrix. In literature a matrix or a problem set defined by \mathbf{A} is said to be ill-conditioned if $\|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ is large [KL80]. Consequently, the condition number of matrix \mathbf{A} is defined as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2, \quad (4.11)$$

where $\|\mathbf{A}\|_2 = \sqrt{\lambda_{max}(\mathbf{A}^H \mathbf{A})}$ is defined as the spectral norm [KL80], λ_{max} denotes the highest eigenvalue of a matrix and ^H denotes the conjugate transpose. By substituting $\sqrt{\lambda_{max}(\mathbf{A}^H \mathbf{A})} = \sigma_{max}(\mathbf{A})$ into eq. (4.11) the condition number can be defined by

$$\kappa(\mathbf{A}) = \frac{\sigma_1(\mathbf{A})}{\sigma_N(\mathbf{A})}, \quad (4.12)$$

where $\sigma_1(\mathbf{A})$ and $\sigma_N(\mathbf{A})$ are the maximal and minimal singular values of \mathbf{A} [CK07]. It can be seen in fig. 4.5(a) that the condition number of the system matrix exhibits high values especially in the lower frequency range.

To overcome this shortcoming the matrix is regularized using the singular value decomposition.

¹Let us assume a linear system $\mathbf{A} \mathbf{x} = \mathbf{b}$. This system is considered well-conditioned if a small change in the system matrix \mathbf{A} or in the right hand side \mathbf{b} results in a small change of the solution vector \mathbf{x} . Vice versa a system is ill-conditioned if small changes or small errors in measurement data result in large changes of the solution vector.

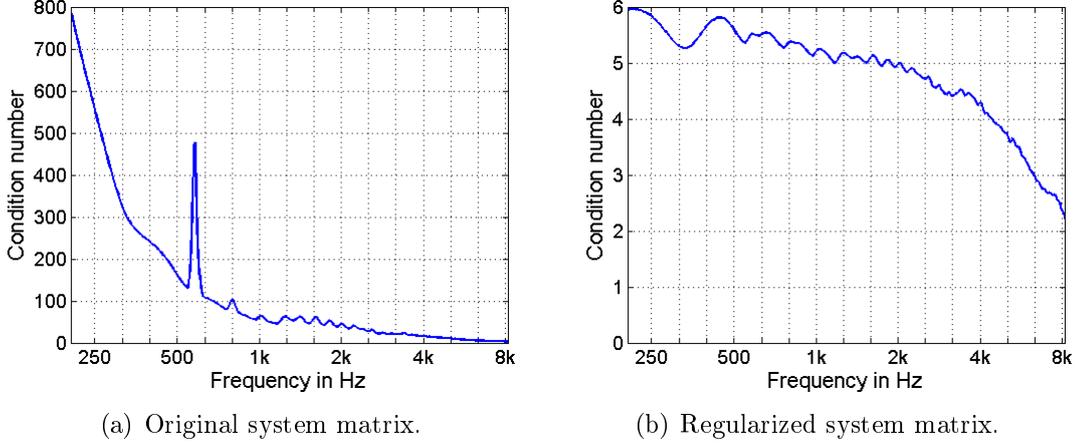


Figure 4.5: Condition number of the system matrix $\hat{\mathbf{H}}$ in the array achievable subspace.

Singular value decomposition (SVD): Let us again consider an arbitrary $N \times M$ rectangular matrix \mathbf{A} where the SVD [JW91], [CK07] is formally defined as the factorization

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^H, \quad (4.13)$$

where \mathbf{U} and \mathbf{V} are unitary matrices (a matrix \mathbf{U} is unitary iff its conjugate transpose equals its inverse $\mathbf{U}^H \mathbf{U} = \mathbf{U} \mathbf{U}^H = \mathbf{I}$) with the left and right singular vectors² of \mathbf{A} in its columns and where \mathbf{S} is a diagonal matrix containing the singular values of \mathbf{A} in descending order

$$\mathbf{S} = \text{diag}_{N \times M}(\boldsymbol{\sigma}), \quad \boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_N]^T \text{ with } \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N. \quad (4.14)$$

Pseudo inverse: As the matrices \mathbf{U} and \mathbf{V} are unitary it is straightforward to define the inverse of the factorized matrix $\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$. After multiplying eq. (4.13) with \mathbf{U}^H from the left and \mathbf{V} from the right we obtain the diagonalised expression of \mathbf{A} as $\mathbf{U}^H \mathbf{A} \mathbf{V} = \mathbf{S}$. The generalized inversion of \mathbf{S} is then defined by

$$\mathbf{S}^+ = \text{diag}_{M \times N} \left(\left[\begin{array}{c} 1 \\ \sigma_i \end{array} \right]_i \right) = \mathbf{V}^H \mathbf{A}^+ \mathbf{U}. \quad (4.15)$$

² The left singular vectors of \mathbf{A} are the orthonormal eigenvectors of $\mathbf{A} \mathbf{A}^H$ whereas the right singular vectors of \mathbf{A} are the orthonormal eigenvectors of $\mathbf{A}^H \mathbf{A}$

Consequently, the generalized inverse of matrix \mathbf{A} is defined as

$$\mathbf{A}^+ = \mathbf{V}\mathbf{S}^+\mathbf{U}^H. \quad (4.16)$$

Further, a regularized inversion is achieved by substituting $\tilde{\mathbf{S}}^+ = \text{diag}_{M \times N} \left(\left[\frac{1}{\sigma_i + \epsilon} \right]_i \right)$, where ϵ is a regularisation variable, into eq. (4.16). A regularization is used to improve the conditioning of the inversion in scenarios where $\sigma_1 \gg \sigma_N$.

System inversion: In the following paragraph we take a detailed look into the regularization of the system matrix $\hat{\mathbf{H}}$ and we examine the effects of two different types of regularizations that will be combined. According to eq. (4.12) the conditioning of $\hat{\mathbf{H}}$ can be improved by changing the ratio of the maximal σ_1 and minimal singular values σ_N . In order to do so we introduce a dynamic range limitation of the singular values

$$\tilde{\sigma} = \sigma + \underbrace{\sigma_1 c_1}_{\text{local}} + \underbrace{\sigma_{max} c_2}_{\text{global}}, \quad (4.17)$$

where $\sigma_{max} = \max_f(\sigma_1(f))$ refers to the maximal singular value over the entire frequency range (global), σ_1 denotes the highest singular value of the system matrix at frequency f (local) and the scalar regularization constants c_1, c_2 are used to control the amount of regularization. The effect of the local and global regularization is explained next

- local regularization: the main task of the local regularization is the improvement of the system conditioning and subsequently, the inversion accuracy.
- global regularization: is used to avoid an extreme amplification of the inverse system at frequencies where one or even all components are not available in the system matrix, $1 \gg \sigma_1 \gg \sigma_N$. The main task of the global regularization is to upwardly limit the filter responses (inverse of the system).

Figure 4.6(a) shows the evolution of the global and local regularization. It can be seen that the global regularization is essential in the lower frequency range whereas above 200 Hz local and global regularization affect the approximation to the same extent. This indicates that the singular values are generally smaller in the lower frequency range of the system as also observed when considering fig. 4.6(b). Additionally, one can recognize that the regularization affects the smaller singular values σ_N relatively

more when compared to σ_1 . Consequently, this means that the most approximation error is introduced to weak components in the system, where the singular values are close to zero, c.f. lower frequency range in fig. 4.6(b).

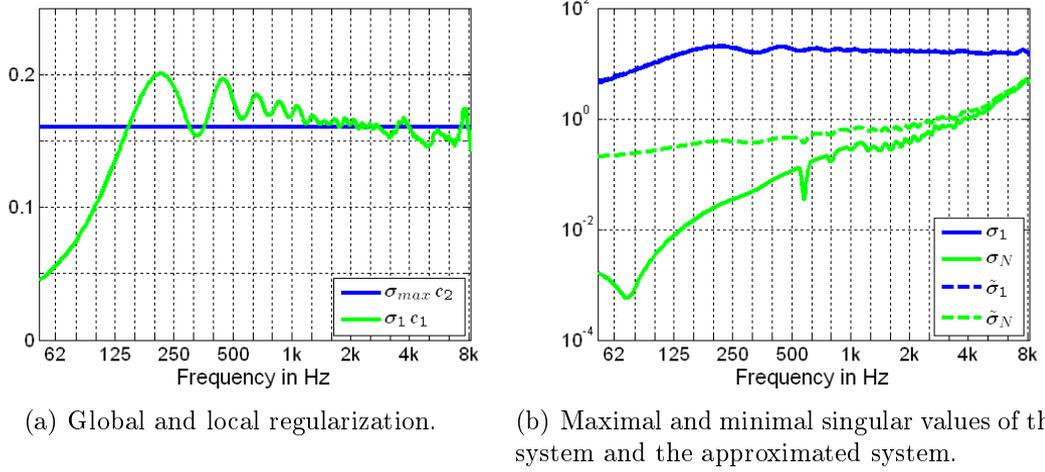


Figure 4.6: Effects of the matrix regularization for $c_2 = 0.008$ and $c_1 = 0.01$.

However, the approximated regularized system matrix for one frequency is defined by

$$\tilde{\mathbf{H}} = \mathbf{U} \tilde{\mathbf{S}} \mathbf{V}, \quad (4.18)$$

where $\tilde{\mathbf{S}}$ denotes the diagonal matrix that contains the regulated singular values defined in (4.17).

The condition number of the approximated system matrix is defined to

$$\tilde{\kappa} = \frac{\tilde{\sigma}_1}{\tilde{\sigma}_N} = \frac{\sigma_1 + \sigma_1 c_1 + \sigma_{max} c_2}{\sigma_N + \sigma_1 c_1 + \sigma_{max} c_2} \quad (4.19)$$

and the maximal condition number is limited to

$$\tilde{\kappa}_{max} = \tilde{\kappa}|_{\sigma_1 = \sigma_{max}, \sigma_N \rightarrow 0} \quad (4.20)$$

$$= \frac{(1 + c_1 + c_2)}{(c_1 + c_2)}, \quad (4.21)$$

which means that it is essentially determined by the selection of the regularization constants c_1 and c_2 . Fig. 4.5(b) illustrates the improved conditioning of the approximated system matrix after a regularization with $c_2 = c_1 = 0.1$.

As defined in eq. (4.16), the actual inversion of the approximated system (4.18) is defined as

$$\tilde{\mathbf{R}} = \mathbf{V} \tilde{\mathbf{S}}^+ \mathbf{U}^H. \quad (4.22)$$

Comparison of inversion and regularized inversion: According to fig. 4.6 the area of greatest alteration is found in the lower frequency segment. Further, it can be seen that the relative change of the minimal singular value is slowly decreasing towards the higher frequency range and reaches at about 5 kHz a point where the approximated system essentially equals the original system. Consequently, we expect the most change caused by the regularization to affect the frequency range below 5 kHz.

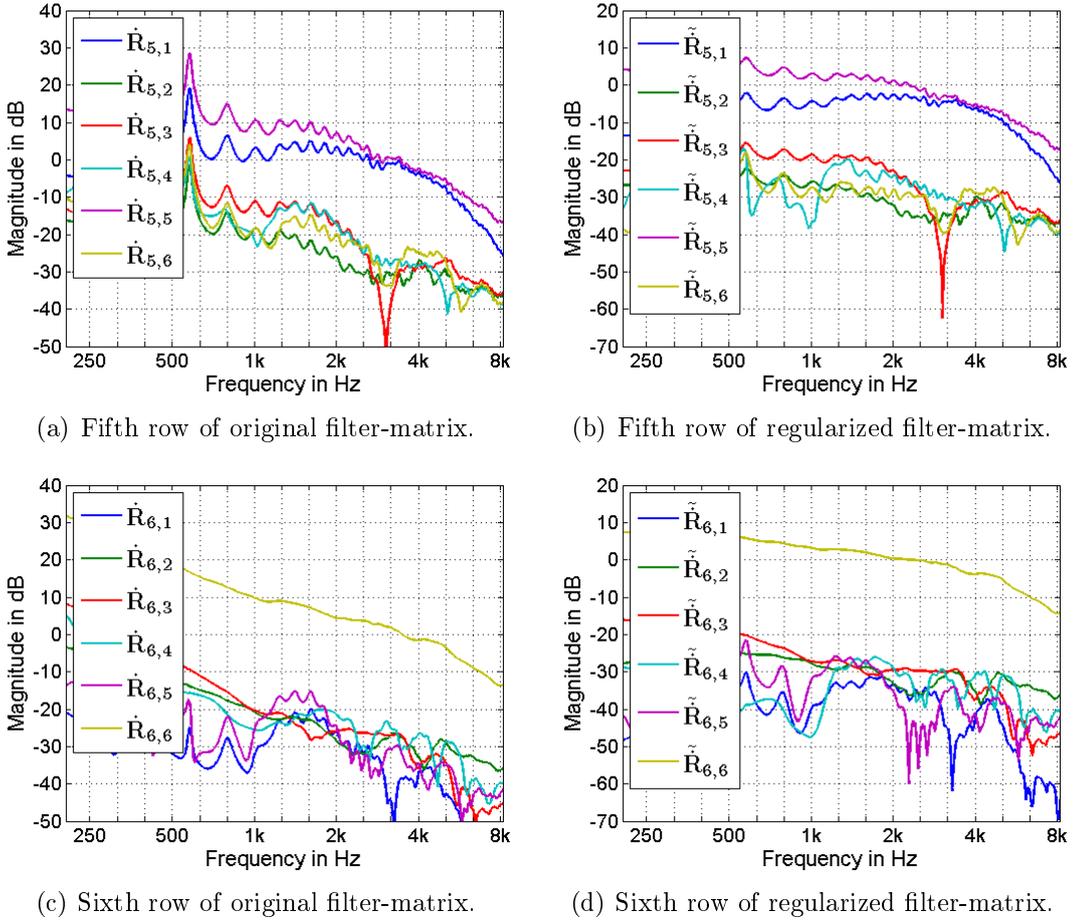
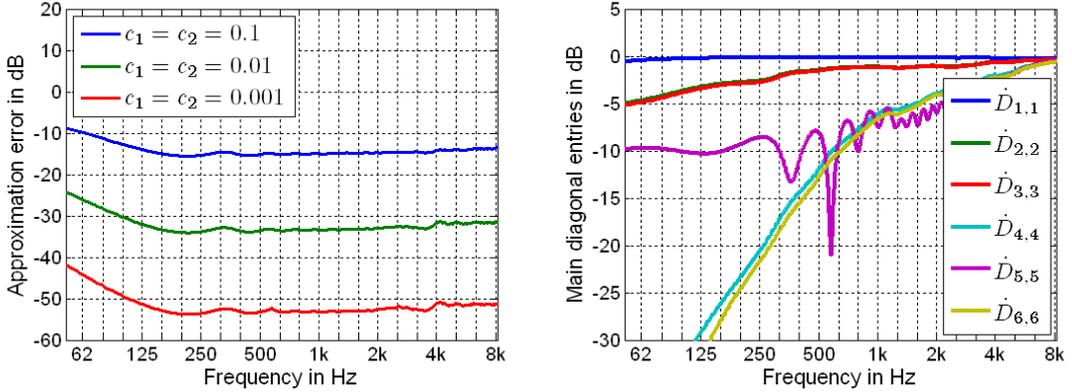


Figure 4.7: Comparison of original and regularized filters for $c_2 = c_1 = 0.01$.

The mentioned effect is further illustrated by fig. 4.7. When considering for example fig. 4.7(c) and fig. 4.7(d) it is apparent that especially the originally high amplifications for high orders at low frequencies are attenuated after the regularization. Another desirable effect is illustrated in fig. 4.7(b) where it is shown that the ripple present for strong components is reduced and thus the magnitude response of the regularized decomposition filters is smoother when compared to the non-regularized filters 4.7(a).

Approximation error: As mentioned, ideally the product of the filter matrix and the system matrix should result in an identity matrix $\tilde{\mathbf{R}}\dot{\mathbf{H}} = \mathbf{I}$. An error measure, which indicates the distance to a identity matrix is defined as $\|\tilde{\mathbf{R}}\dot{\mathbf{H}} - \mathbf{I}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. Again note that the frequency dependence is dropped to improve readability. The evolution of the error over frequency is shown in fig. 4.8(a) where one can observe that the approximation error is essentially flat over frequency and obviously depends on the settings of c_1 and c_2 .

Another interesting figure is presented in 4.8(b) where one can study the evolution of the absolute values on the main diagonal of the overall system matrix $\dot{D} = \tilde{\mathbf{R}}\dot{\mathbf{H}}$. Accordingly, the modes of higher order are not available for low frequencies, as the regularization limits the amplification.



(a) Approximation error.

(b) Evolution of main diagonal entries according to regularization with $c_1 = c_2 = 0.01$.

Figure 4.8: Effects of regularization.

4.1.2 Summary

In this section we described a measurement data based spherical harmonics decomposition of the microphone signals \mathbf{x} . It was shown that the decomposition into the orthogonal spherical harmonics spectrum $\hat{\phi}_{N_A}$ requires a 6×6 decomposition matrix. Figure 4.9 represents the two main stages of the decomposition process consisting of an encoder stage that provides frequency dependent and modal aliased modes and a filter matrix that is designed to separate the aliased modes and unify their response over frequency.

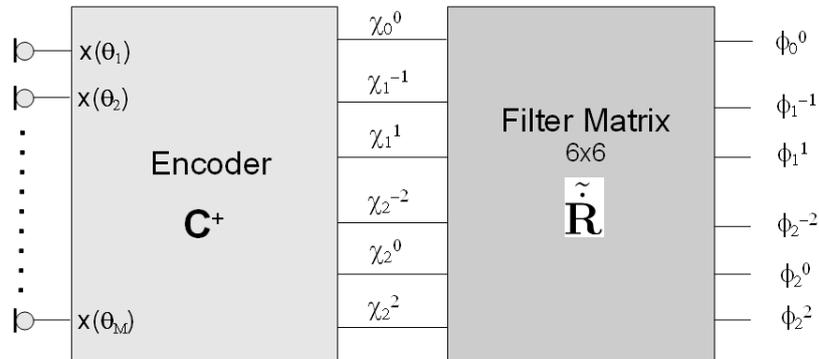


Figure 4.9: Scheme of decomposition and filtering.

4.2 Beamforming

A block diagram of the beamformer unit is depicted in fig. 4.10. Thereby, the input signals are the unified modes in the spherical harmonics domain ϕ_n^m that are generated by the decomposer unit. In the steering unit these signals are weighted with the spherical harmonics evaluated at the lookdirection θ_1 and in a next step they are multiplied with frequency independent order weights w_n that are designed to form specific beam-pattern shapes (see sections 2.1 and table 2.3). In the last unit, the summation unit, the obtained signals are summed up and normalized such that the beam amplitude in lookdirection is constant over all steering directions and beampattern shapes.

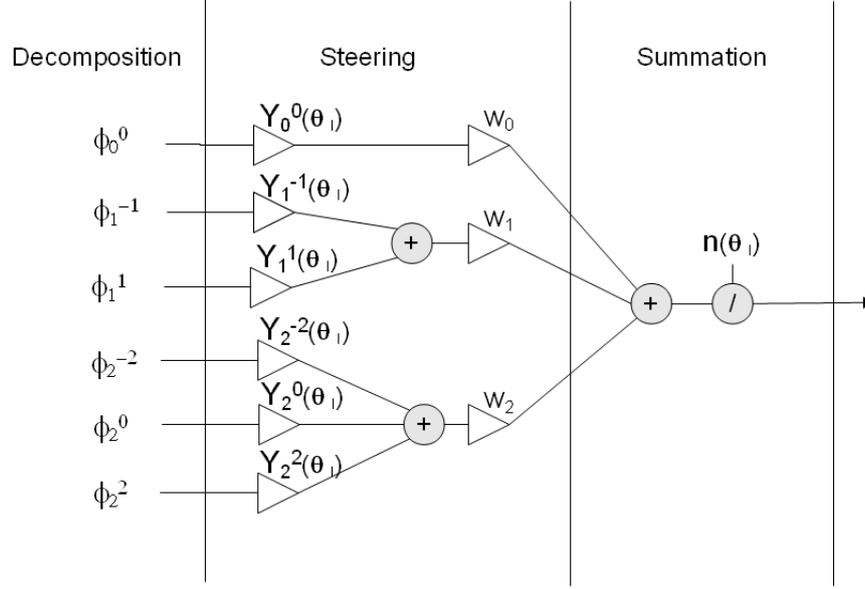


Figure 4.10: Scheme of beamforming.

4.2.1 White noise gain (WNG)

The white noise gain is a robustness measure with respect to errors in the array setup [EKM04].

Generally the WNG is defined as the ratio of the signal energy at the output of the array and the energy of the measurement noise at the array output

$$WNG(kr) = \frac{|d_s(kr)|^2}{E\{|d_e(kr)|^2\}}, \quad (4.23)$$

where d_s denotes the signal energy and d_e the noise energy at the output of a beamforming system. As it is common to assume a plane-wave sound field and an array looking towards the plane-wave direction when analyzing the array robustness, it is possible to define the weighting coefficients and the sound-field coefficients as

$$\begin{aligned} \chi_n^m(kr) &= 4\pi i^n \rho(kr) Y_n^m(\boldsymbol{\theta}_1) \\ g_n^m(kr) &= \frac{w_n}{4\pi i^n \rho(kr)} Y_n^m(\boldsymbol{\theta}_1). \end{aligned} \quad (4.24)$$

Accordingly the signal energy of a unit amplitude plane-wave and the output noise energy of an assumed spatially uncorrelated white noise with unit variance is obtained

by substituting eq. (4.24) into eq. (2.42). The signal energy is then defined by

$$|d_s(k\mathbf{r})|^2 = \left| \sum_{n=0}^N w_n \sum_{l=0}^n Y_n^{(2l-n)}(\boldsymbol{\theta}_1) Y_n^{(2l-n)}(\boldsymbol{\theta}_1) \right|^2 \quad (4.25)$$

and the noise energy by

$$E \{ |d_e k\mathbf{r}|^2 \} = \sum_{j=0}^{M-1} \alpha_j^2 \left| \sum_{n=0}^N \sum_{l=0}^n w_n \frac{1}{i^n 4\pi \rho_n(kr_j)} Y_n^{(2l-n)}(\boldsymbol{\theta}_1) Y_n^{(2l-n)}(\boldsymbol{\theta}_j) \right|^2, \quad (4.26)$$

where α_j^2 is a discretized integration constant (eq. (A.33)) and the computation is defined for the considered case of hemispherical beamforming, see sec.2.2.4. For a computation of the WNG of the presented beamforming-system, the compensation of $i^n 4\pi \rho_n(kr_j)$ is substituted by the corresponding entry of the decomposition matrix.

Note that the output energy of a beamformer where just the even symmetric spherical harmonics are considered varies over different zenith angles. Therefore the resulting WNG is depicted for special zenith angles of $\boldsymbol{\theta}_1$, see fig. 4.11.

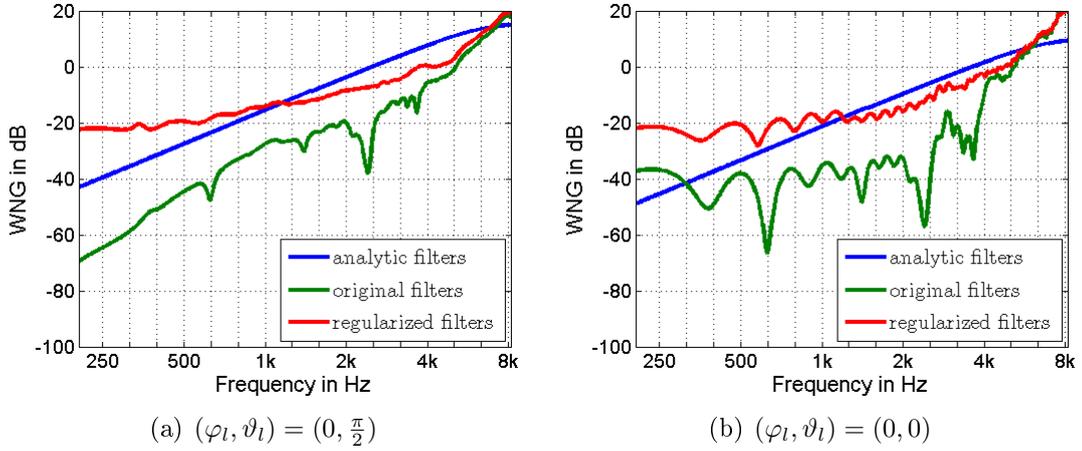


Figure 4.11: WNG of second order beamforming system with hypercardioid weighting.

As the computation of the WNG includes the frequency response compensation $\rho_n(kr)$ it is evident that at low frequencies the high orders will have main impact on the calculation (cf. fig. 4.4(b) and fig. 5.1(b)). For $(\varphi_l, \vartheta_l) = (0, \frac{\pi}{2})$ (fig. 4.11(a)) the output energy is mainly dependent on the chronologically first, third and sixth harmonic and accordingly, the frequency response compensation that is needed to

unify the the sixth mode, strongly contributes to the WNG at low frequencies (cf. fig. 4.4(d)). On the other hand, for $(\varphi_l, \vartheta_l) = (0, 0)$ (fig. 4.11(b)) the WNG at low frequencies strongly depends on the frequency compensation of the fifth harmonic. Further it can be seen in fig. 4.11 that the regularization of the compensation filters improves the WNG in the lower frequency regions.

The WNG can be seen as a measure that represents the improvement of the signal to noise ratio at the output compared to the input or as a general measure of robustness [Raf05]. Accordingly, the beamforming system lacks robustness especially in the lower frequency regions and for non-regularized decomposition filters, see fig 4.11. In order to minimize the influence of the decomposition filters for higher orders on the WNG, the obvious solution would be to implement beampatterns of lower order. Unfortunately, this would bring up the disadvantage of wasting directivity at higher frequencies.

Implementation strategies that combine the above considerations design beampatterns for a constant WNG and maximize the directivity index (DI). The obtained order weights are then obviously frequency dependent $w_n(k)$ [EKM03]. Other optimization methods consider the design of the beamforming weights as a multiply constrained problem, such that a resulting beamformer offer a trade-off among various conflicting performance measures such as the DI, WNG, array gain, mainlobe width and so on [YSS⁺11]. In general those strategies result in beampatterns of lower order for low frequencies and increasing pattern-order as the frequency increases. The transitions between different orders can be smoothed according to [BPF11].

As the regularization of the decomposition filters already improves the robustness as can be seen in fig. 4.11, we will not explain the constrained weighting coefficients design in greater detail and instead focus on the implementation of patterns with frequency invariant weighting coefficients w_n , see chapter 2.2. However, the design of frequency variant weighting coefficients may be subject to future work.

4.3 Implementation and practical aspects

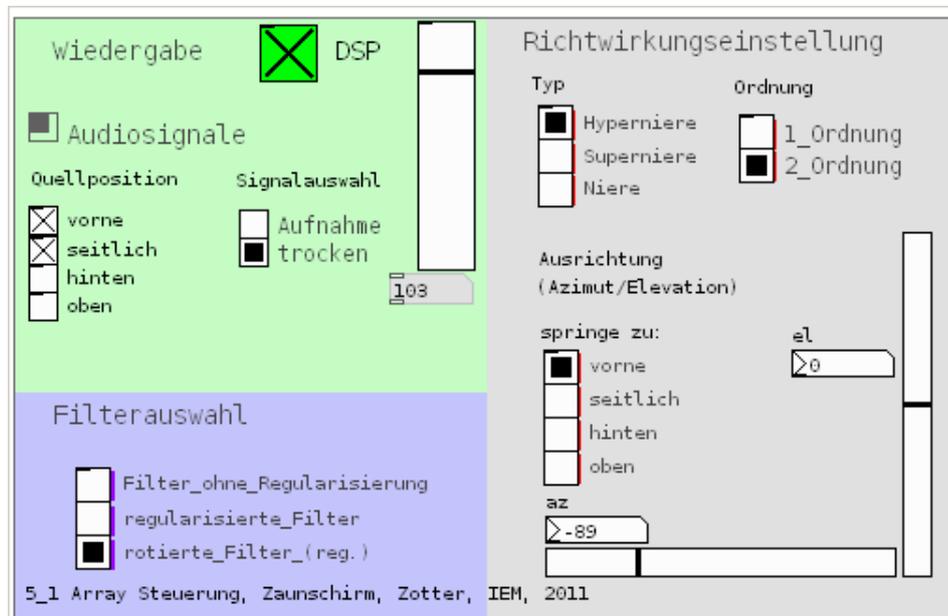
In order to test the functionality and the quality of the presented second order beamforming approach we implemented a patch in pure data (pd) that combines all above considerations into a working real-time system.

The time-varying audio signals that are recorded by the sensors of the microphone array are decomposed into time-varying series of signals in the spherical harmonics domain. In a next step these signals are filtered with a 6×6 filter matrix such that aliased modes are separated (orthogonalized) and their frequency response is unified such that each each mode is available over the entire frequency range.

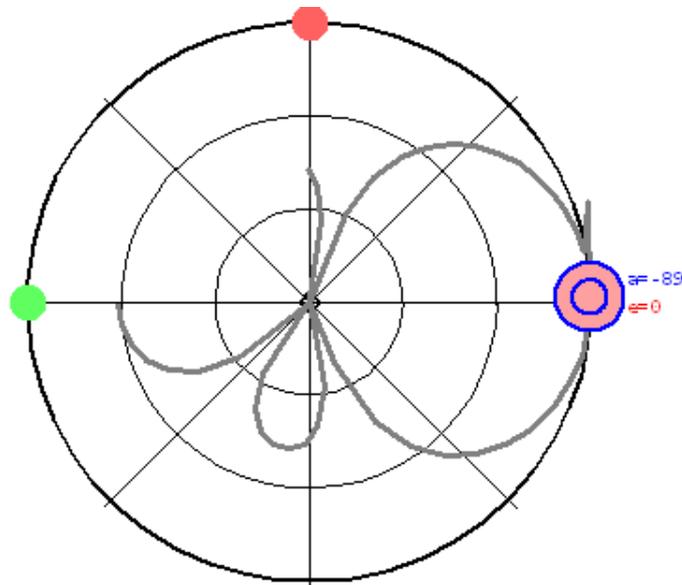
The actual filtering is done by a convolution of the transformed microphone signals with the filter impulse responses. This convolution is implemented as a partitioned overlap-and-save algorithm [TF01] in pd (`partconv`) with a latency defined as partition size minus the intern block size. In the last step the unified signals are steered and combined as explained in section 4.2.

For the purpose of testing the proposed approach we generated audio samples using the measured IRs between a specific speaker and each microphone button. Additionally, samples were directly recorded in order to simulate the practical usage as a for example a teleconferencing microphone.

In order to allow for a simple and efficient testing and comparison of different beamforming orders (switch between 1st and 2nd order beamforming) as well as weighting and filtering approaches we implemented a graphical user interface (see fig. 4.12(a)) that allows for switching of all aforementioned parameters. As an additional feature, the program indicates the look direction, the active sources and the current directivity pattern in an additional window. An example is given in fig. 4.12(b).



(a) Graphical user interface.



(b) Source and look direction indication.

Figure 4.12: Graphical user interface and positioning indicator.

4.3.1 Generated beampatterns

Due to the regularization of the decomposition filters presented in section 4.1.1 it obvious that the practically implemented beampatterns differ from the ideal beampatterns depicted in figures 2.4 to 2.6 and 2.11.

The achievable beampatterns with a supercardioid weighting steered towards the horizontal array plane are shown in fig. 4.13. It is striking that the characteristic are

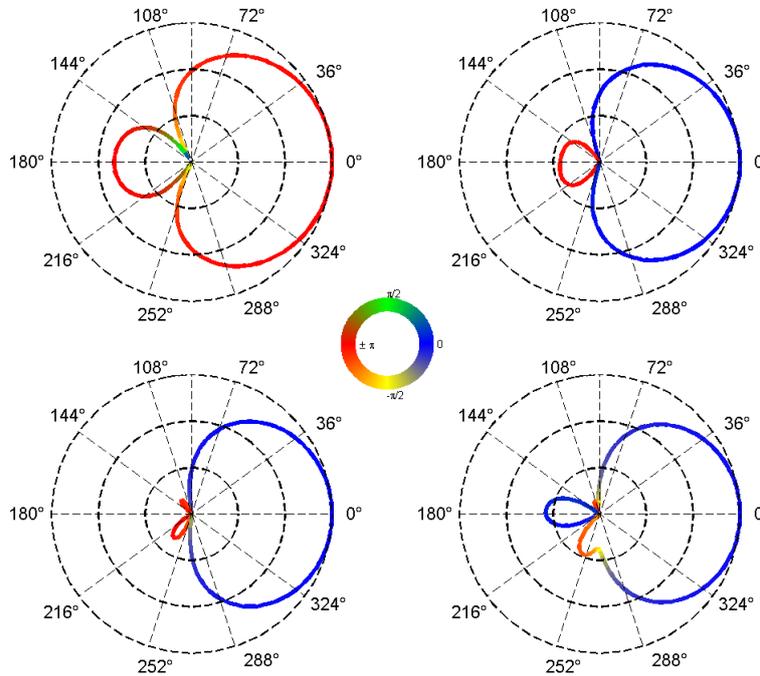


Figure 4.13: Supercardioid beampatterns at frequencies of $[400, 1000, 2000, 4000]Hz$ and usage of regularized filters ($-40dB$).

frequency dependent and that the beampatterns evolve from a nearly first order supercardioid at low frequencies to a second order supercardioid characteristic at high frequencies. The vertical slices of a beam steered towards the z-direction are shown in fig. 4.14 where one can observe a similar behaviour as for the horizontal slices, namely that the higher order pattern is just available at higher frequencies. The same effect is observed for the other types and therefore they are not considered separately. The quasi order limited beamforming is caused by the filter regularization, as the higher

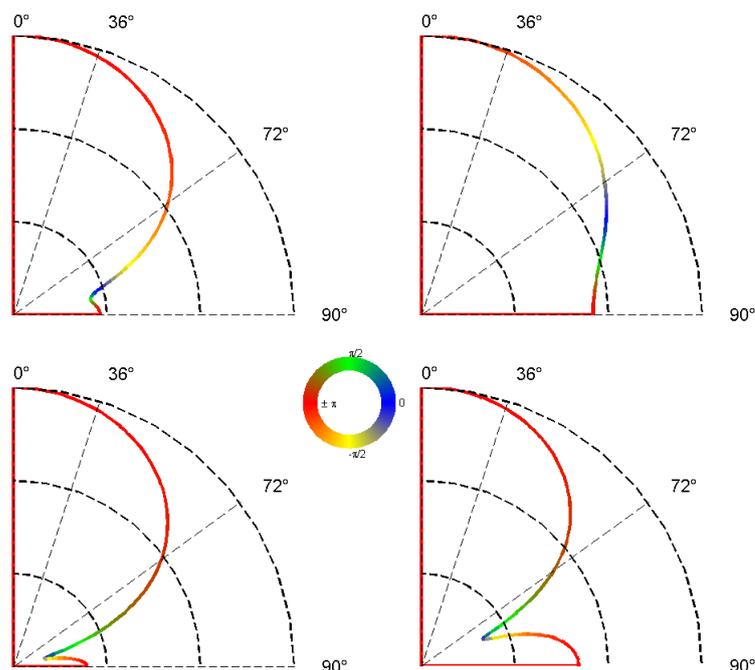


Figure 4.14: Hypercardioid vertical beampatterns at frequencies of $[400, 1000, 2000, 4000]Hz$ and usage of regularized filters ($-40dB$ regularization).

order modes are attenuated in the lower frequency range, see fig. 4.8(b). Additionally, this order limitation improves the WNG and accordingly the robustness of the beamforming system [EKM⁺09].

In order to test the portability of the designed decomposition filters to an array duplicate with same geometry but other not perfectly matched cardioid microphones we generated filters with the circularly rotated data of the measured matrix H . In a next step the beamforming-system was tested with the rotated filters and the original microphone array. This setup quasi simulates a operating of the developed approach with a modified array. As can be seen in fig. 4.15 the produced beams of this setup highly vary from the beams produced with the original decomposition filters. Thus, the designed filters are not necessarily applicable to arrays where the microphone characteristics differ from the ones mounted on the prototype.

Since it is shown that the designed filters are not applicable to other arrays of the same

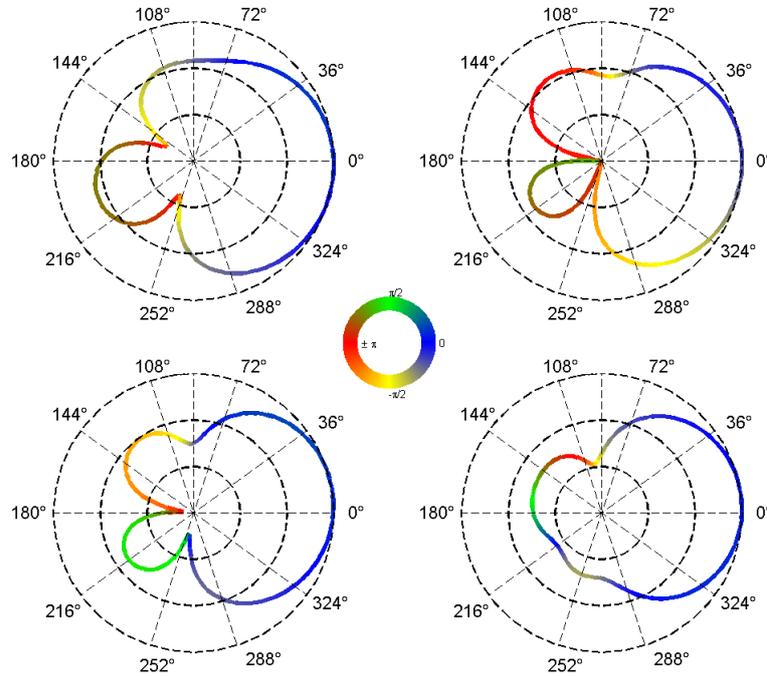


Figure 4.15: Supercardioid beampatterns at frequencies of $[400, 1000, 2000, 4000]Hz$ and usage of rotated filters ($-40dB$).

type but with altered microphone characteristics it is anyhow necessary to measure each array separately in order to design adequate separation filters. Therefore, the complexity needs to be reduced to allow for possible series production. This is achieved by reducing the number of measurement points drastically from previously 288 to 6, whereby each cardioid button is measured on its main axis at a zenith angle of $\vartheta = 84^\circ$ and an additional measurement point on the latitude circle near the pole is selected. When comparing the practical available directivity patterns that are produced with the reduced-complexity filters (presented in figures 4.16 and 4.17) with the directivity patterns achieved by the original filters one can easily observe that the results are comparable. This leads to the conclusion that the measurement complexity can be extremely reduced without degrading the performance.

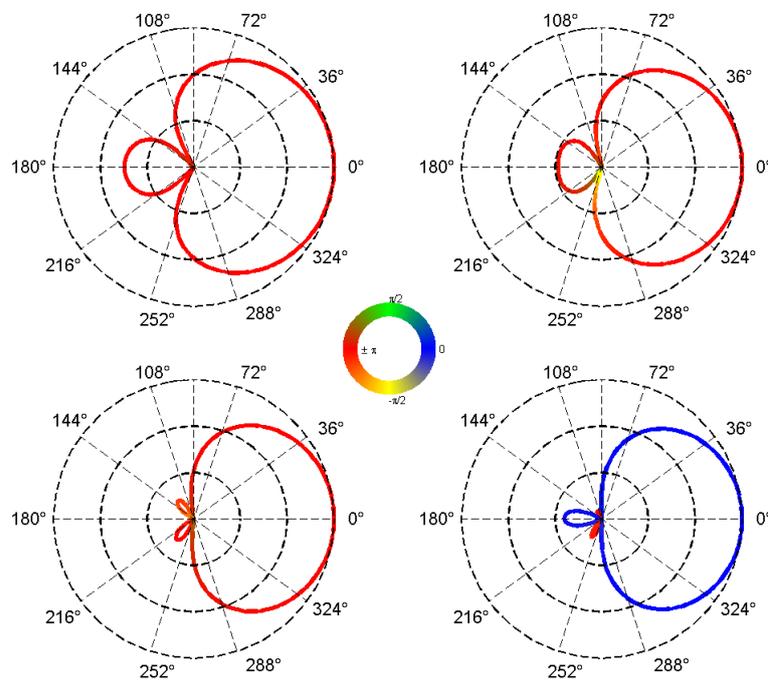


Figure 4.16: Supercardioid beampatterns at frequencies of $[400, 1000, 2000, 4000] Hz$ and usage of reduced-complexity filters ($-40dB$ regularization).

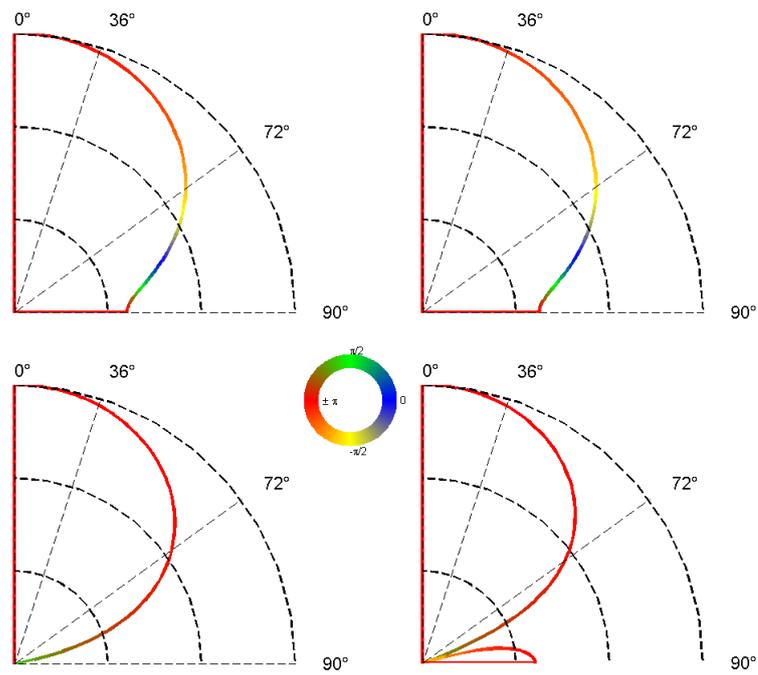


Figure 4.17: Hypercardioid vertical beampatterns at frequencies of $[400, 1000, 2000, 4000] Hz$ and usage of reduced-complexity filters ($-40dB$ regularization).

5. Analytic model of circular microphone arrays

This section deals with the analytic estimation of the holographic spectrum b_n^m out of the sensor signals x of a circular aperture or microphone array.

All derivations are done in the spherical coordinate system, see eq. (A.1) and fig. A.1.

Under the assumption that the circular aperture or array is placed on a boundary surface in the horizontal plane of the coordinate system $\boldsymbol{\theta}|_{\vartheta=\pi/2}$ the sound-field at the aperture is defined as

$$x(kr, \varphi, \pi/2) = \sum_{n=0}^{\infty} \sum_{l=0}^n \chi_n^{(2l-n)}(kr) Y_n^{(2l-n)}(\varphi, \pi/2), \quad (5.1)$$

where $(2l - n)$ is used to define the spherical harmonics degree m , see sec. 2.2.4.

5.1 Continuous circular aperture

Circular harmonics transform: Any arbitrary circular function or pattern can be represented by a spatial Fourier series expansion. With a circular function $\eta(kr, \varphi)$ at aperture radius r , the 'circular harmonic coefficients' η_m and the 'circular harmonics' $\Phi_m(\varphi)$, the transformation pair is defined to

$$CHT \{ \eta(kr, \varphi) \} = \eta_m(kr) = \int_0^{2\pi} \eta(kr, \varphi) \Phi_m(\varphi) d\varphi \quad (5.2)$$

$$ICHT \{ \eta_m(kr) \} = \eta(kr, \varphi) = \sum_{m=-\infty}^{\infty} \eta_m(kr) \Phi_m(\varphi) d\varphi. \quad (5.3)$$

Similar to the spherical harmonics definition (see eq. (A.12)), these circular harmonics can be defined as real functions:

$$\Phi_m(\varphi) = \frac{1}{\sqrt{\pi}} \begin{cases} 1/\sqrt{2}, & m = 0 \\ \cos(m\varphi), & m > 0, \\ \sin(m\varphi), & m < 0 \end{cases} \quad (5.4)$$

where the orthonormality condition:

$$\int_0^{2\pi} \Phi_m(\varphi)\Phi_{m'}(\varphi) d\varphi = \delta_{mm'} \quad (5.5)$$

must hold.

Analysing at the aperture: When separating the azimuth and zenith dependent parts of the spherical harmonics (see eq. (A.13)) eq. (5.1) is rewritten to

$$x(kr, \varphi, \pi/2) = \sum_{n=0}^{\infty} \sum_{l=0}^n \chi_n^{(2l-n)} N_n^{|(2l-n)|} P_n^{|(2l-n)|}(\cos(\pi/2)) \Phi_{(2l-n)}(\varphi), \quad (5.6)$$

where $\Phi_m(\varphi)$ denotes the above described orthonormal functions according to the azimuth and $N_n^{|m|} P_n^{|m|}(\cos(\vartheta))$ denotes orthonormal functions according to the zenith (note that $N_n^{|m|}$ is different from the general spherical harmonics definition (A.17) as the azimuthal functions are normalized separately)

$$\int_{-1}^1 N_n^{|m|} P_n^{|m|}(\cos(\vartheta)) N_{n'}^{|m|} P_{n'}^{|m|}(\cos(\vartheta)) d\cos(\vartheta) = \delta_{nn'},$$

$$\text{with } N_n^m = \sqrt{\frac{(2n+1)(n-m)!}{2(n+m)!}}. \quad (5.7)$$

By substituting eq. (5.6) into eq. (5.2) and with the orthonormality condition in eq. (5.5) the circular harmonic coefficient of degree m is determined by

$$\eta_m(kr) = \sum_{l=0}^{\infty} \chi_{(2l+|m|)}^m(kr) N_{(2l+|m|)}^{|m|} P_{(2l+|m|)}^{|m|}(\cos(\pi/2)), \quad (5.8)$$

where the spherical wave spectrum χ_n^m can be expressed as the product of the holographic spectrum b_n^m and a radial and frequency dependent term $\rho_n(kr)$, see (2.33) and appendix A.4.

5.2 Circular arrays

5.2.1 Sampling the circle

In practice, the continuous aperture is sampled by a set of microphones located at discrete positions r on the circle, where $j = 0, \dots, M - 1$ denotes the j^{th} microphone on the concentric circle with radius r .

For a circular array designed to resolve the sound-field up to a maximal angular spatial frequency N the minimum number of required microphones is $M \geq 2N + 1$. Thus, the minimal number of microphones for a second order array is five, see array prototype fig. 3.1.

With the assumption of a perfectly band limited wave field, where N is the maximum present angular frequency, the discretized transform is spatial-aliasing free. If the wave field is not considered perfectly band limited (higher angular frequencies are present), then the aliasing free decomposition is dependent on the array radius r and the largest wavenumber k_{max} that is present in the sound-field. The upper limiting frequency of the essentially aliasing free frequency-band is defined by

$$f_u = \frac{N c}{2\pi r}, \quad (5.9)$$

where N denotes the maximal resolvable angular frequency of the microphone array, [VT02], [Raf05], [SEH08], [WA01],[Teu06]. This definition is based on the observation that the radial functions $\rho_n(kr)$ (see eq. (2.33)) for higher orders ($> N$) exhibit negligible values for arguments smaller than $k_{max}r$ (cf. figures 5.1(a) and 5.1(b)). Consequently, these orders are sufficiently attenuated for wave-numbers below k_{max} .

Substituting $M > 2N$ into equation (5.9) also gives indication of the relation between the sampling spacing on the circle and the smallest wavelength λ_{min} that is present in the wave field

$$d_{sampling} \leq \lambda_{min}/2. \quad (5.10)$$

In other words this means that spherical aliasing becomes critical when the sample spacing exceeds half of the acoustic wavelength.

However, the equiangular sampling interval of a N^{th} order circular array is defined by

$$\Delta\varphi = 2\pi/(2N + 1). \quad (5.11)$$

With the above considerations a second order circular microphone has to consist of at least five microphones, that are positioned at $\varphi_j = (\frac{360^\circ}{5})j$. And with an array radius of $r = 0.021\text{m}$ the quasi spatial aliasing free analysis is limited to $f_u = 5.2 \text{ kHz}$.

Discretized functions in azimuth: For the chosen equally spaced sampling it is ensured that the basis functions $\Phi_m(\varphi)$ remain orthonormal with a discretized integration constant $\Delta\varphi = \frac{2\pi}{2N+1}$

$$\frac{2\pi}{2N + 1} \sum_{j=0}^{2N} \Phi_m(\varphi_j) \Phi_{m'}(\varphi_j) = \delta_{m'm} \quad (5.12)$$

and it is possible to transform the recorded microphone signals x into the circular harmonics domain.

Further, let us express the microphone signals in terms of a spherical harmonics expansion

$$x_j(\varphi_j) = \sum_{n=0}^N \sum_{l=0}^n b_n^{(2l-n)} \rho_n(kr) N_n^{(2l-n)} P_n^{(2l-n)}(0) \Phi_{(2l-n)}(\varphi_j), \quad (5.13)$$

where the radial function $\rho_n(kr)$ is defined dependent on the used microphone type and $(2l - n)$ is used to define the spherical harmonics degree.

In order to obtain the different circular harmonic modes $\eta_m(kr)$, the $(2N + 1)$ microphone signals are weighted with $\frac{2\pi}{2N+1} \Phi_{m'}(\varphi_j)$ and summed over the index j . Due to the orthonormality of the azimuthal functions the sum over l on the right hand side of the equation vanishes and $\frac{2\pi}{2N+1}$ is chopped

$$\eta_m(kr) = \sum_{j=0}^{2N} x_j \Phi_m(\varphi_j) = \sum_{l=0}^{\lfloor \frac{N-|m|}{2} \rfloor} b_{(2l+|m|)}^m \rho_{(2l+|m|)}(kr) N_{(2l+|m|)}^{|m|} P_{(2l+|m|)}^{|m|}(0). \quad (5.14)$$

5.2.2 Frequency dependence of modes

As is evident from eq. (5.14) the modes are dependent on frequency and aperture or array radius. This dependency is given by the functions $\rho_n(kr)$ and consequently

by the spherical Bessel function and its derivative. The magnitude responses of the decomposed components are depicted in figures 5.1(a) for omnidirectional transducers and 5.1(b) for cardioid microphones.

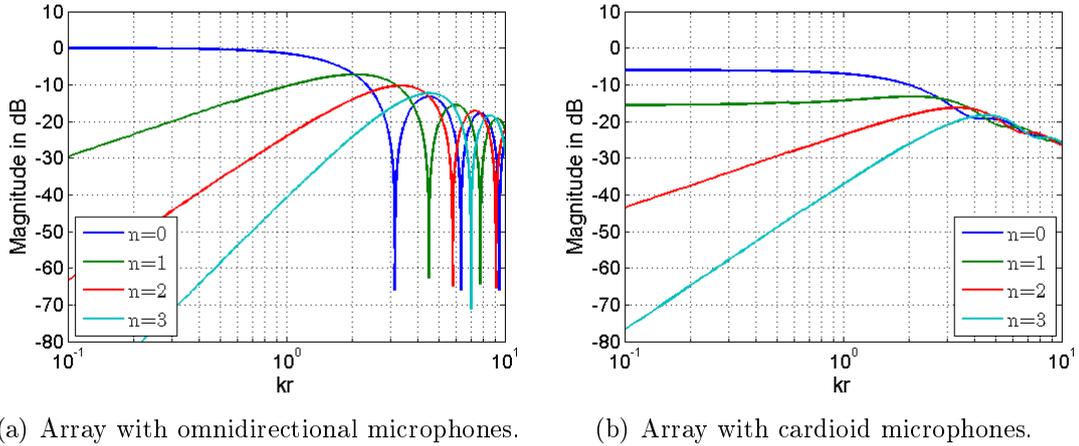


Figure 5.1: Magnitude of $\rho_n(kr)$ up to order $n = 3$.

It can be seen in figure 5.1(a) that the radial function for order $n = 0$ is almost flat over the lower frequency range and that higher orders exhibit a high-pass response of order n (for $n = 1$ the slope is 20dB/decade). In the upper frequency range one can observe notches in the responses of all orders (cf. zeros of the Bessel function). These singularities have negative effect on the availability of the circular harmonic modes at these singularities and complicate or impede an inversion of the function that is necessary to flatten the frequency behaviour of the modes. However, an inversion of $\rho_n(kr)$ is indispensable if the microphone array is used for beamforming and thus it is not applicable to attain a frequency independent desired beampattern. This problem is also well known from open sphere array configurations. Strategies to overcome the mentioned difficulties are for example given in [BR07], [Teu06] and [Raf05] and include the use of directional elements e.g. cardioid microphones instead of omnidirectional microphones and the mounting of the circular aperture or array into a rigid cylindrical baffle. By comparing figure 5.1(b) with figure 5.1(a) one can easily observe the desirable effect that is achieved by using cardioid microphones. The main advantage is the lack of dips in the frequency response (there are no common zeros in the Bessel function and its derivative) which makes the function practically invertible. Further, both the zero-order ($n = 0$) and the first-order ($n = 1$) component

exhibit an essentially flat response over the lower frequency range (simple inversion). The higher orders $n > 1$ exhibit a high-pass response of order $n - 1$ up to $kr \sim 1$ (e.g. the component of order $n = 2$ has a slope of 20dB/decade).

5.2.3 Modal aliasing

Another issue that can be derived from eq. (5.14) is referred to as modal aliasing. With the determined circular modes $\eta_m(kr)$ it is indeed possible to control a beam-pattern in the horizontal plane (array plane) but this approach lacks control over the vertical spatial response.

In order to control the vertical response it is necessary to gain control over the spherical harmonics coefficients (see A.3). From eq. (5.14) it is evident that different spherical harmonic modes contribute to one circular mode, as all spherical harmonic coefficients with the same degree m are combined in the corresponding circular harmonics coefficient $\eta_m(kr)$ (see figure 5.2). This specific issue is referred to as modal aliasing¹. It is seen that the chronologically first modal aliasing occurs for degree $m = 0$. Let us consider a band-limited wave-field ($N = 2$), then the spherical harmonics of order and degree $n = 0, m = 0$ and $n = 2, m = 0$ both contribute to the mode $\eta_0(kr)$. If the wave-field is not band-limited at all, it is evident that all spherical harmonics with the same degree contribute to the corresponding circular mode if they are available at the specific frequency ($N \approx kr$).

5.2.4 Accessing the spherical harmonic modes

The additional information that is presented by a centre element or in general by additional concentric circular arrays is the key to access all spherical harmonics, see [ME08b]. Since the modal aliasing is in general not dependent on microphone type that is used to sample the sound-field on the concentric circle, this approach is applicable for both directional and omnidirectional sensors. The actual separation process is explained for the implemented prototype 3, see fig. 5.3.

¹Note that modal aliasing is present for continuous apertures as well as for discrete sampled arrays and that it is independent of the used microphone type. Accordingly, it is not a result of the sampling scheme or the procedure of sampling in general (cf. spatial aliasing A.5.2)

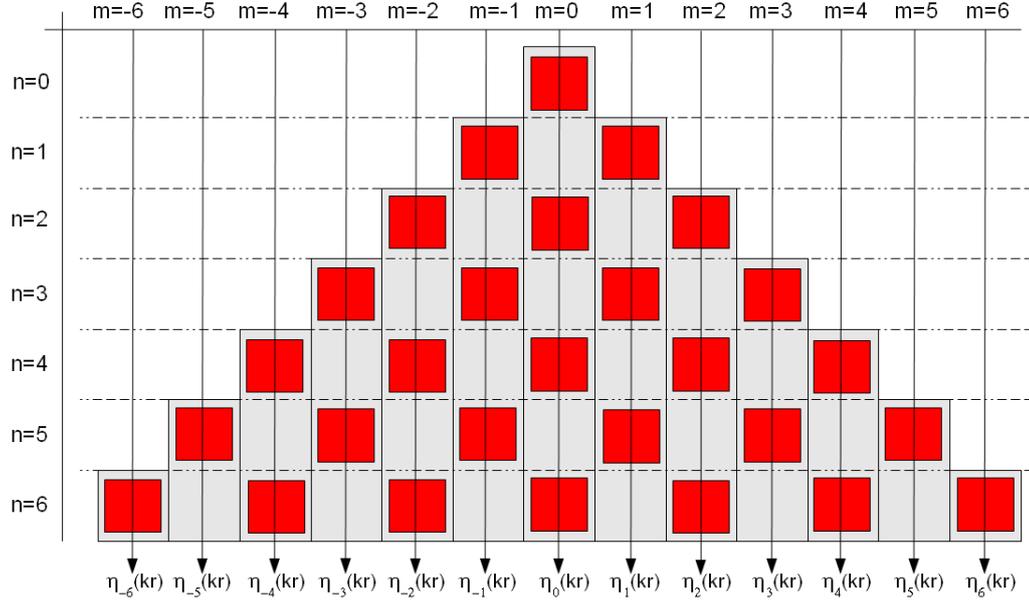


Figure 5.2: Graphic interpretation of modal aliasing up to order $n = 6$.

With the five cardioid microphones at the concentric circle ($\Delta\varphi = 72^\circ$) the sound-field is decomposed into five circular harmonic modes that are defined in eq. (5.14). Under the assumption of a band limited sound-field with maximal angular spatial frequency N the only aliased mode is found for $m = 0$ which is also true for a not band limited sound-field up to $f_u = 5.2$ kHz.

The circular mode for $m = 0$ and with $N_0^{|0|} = 1/\sqrt{2}$, $P_0^{|0|}(0) = 1$, $N_2^{|0|} = \sqrt{5}/\sqrt{2}$ and $P_2^{|0|}(0) = -0.5$ is given as

$$\sum_{j=0}^4 x_j \frac{1}{\sqrt{2\pi}} = \left[b_0^0 (j_0(kr) - i j_0'(kr)) \sqrt{\frac{1}{2}} - b_2^0 (j_2(kr) - i j_2'(kr)) 0.5 \sqrt{\frac{5}{2}} \right]. \quad (5.15)$$

By using the fact that the spherical Bessel functions for argument ($kr = 0$) are zero for all orders $n > 0$ (see fig.5.1(a)), the signal captured by the centred omnidirectional microphone is

$$\dot{x} = b_0^0 j_0(0) \frac{1}{\sqrt{4\pi}}. \quad (5.16)$$

After multiplying (5.15) with $\sqrt{2}$ we obtain a matrix notation of the separation system

$$\begin{bmatrix} \dot{x}\sqrt{4\pi} \\ \sum x_j/\sqrt{\pi} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ j_0(kr) - i j_0'(kr) & -5\sqrt{5}(j_2(kr) - i j_2'(kr)) \end{bmatrix} \begin{bmatrix} b_0^0 \\ b_2^0 \end{bmatrix} \quad (5.17)$$

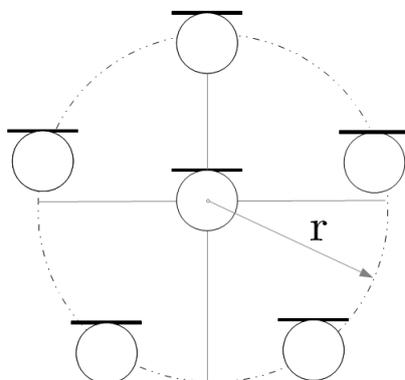


Figure 5.3: Schematic layout of the prototype.

and after inverting the system, the spherical harmonic coefficient of order $n = 2$ and degree $m = 0$ is

$$b_2^0 = \frac{\sum x_j / \sqrt{\pi} - \hat{x} \sqrt{4\pi} ((j_0(kr) - i j_0'(kr)))}{-0.5 \sqrt{5} (j_2(kr) - i j_2'(kr))}. \quad (5.18)$$

The spherical harmonic modes of degree $m \neq 0$ are directly accessed with the circular harmonics $\eta_m(kr)$, see fig. 5.4. In order to unify the responses of the spherical

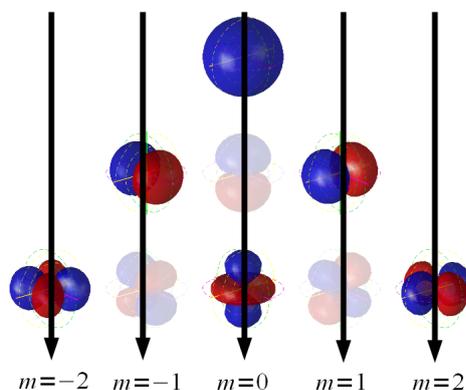


Figure 5.4: Relation between circular modes and spherical-harmonics.

harmonic modes, a frequency-response compensation R_n is employed, cf. A.4 relation between spherical wave spectra and wave spectra. If there is no modal aliasing occurring, these filters are defined straightforward as

$$R_n(kr) = \frac{1}{(j_n(kr) - i j_n'(kr))}, \quad (5.19)$$

where $N_n^{|m|} P_n^{|m|}(0)$ are not included. In literature these filters are referred to as holographic filters, see [Ple09]. With the coefficients of the interior problem b_n^m the incident wave field is completely represented in the spherical harmonics domain. With the radial filters for incident plane waves ($k r_i \geq 1$)

$$\mathbf{P} = \begin{pmatrix} 4\pi & 0 & \dots & 0 \\ 0 & 4\pi i^n & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 4\pi i^N \end{pmatrix} \quad (5.20)$$

the targeted modal source spectrum ϕ is obtained by

$$\phi_N = \mathbf{P}^{-1} \mathbf{b}_N. \quad (5.21)$$

Finally, the separation process is illustrated in fig. 5.5. Summarized, the sound-field is sampled by 5 microphones on a concentric circle and transformed into circular modes, that are filtered in order to obtain the targeted unified spherical harmonic modes. Further, the modal-aliased mode η_0 is separated by using the signal of a centred omnidirectional sensor.

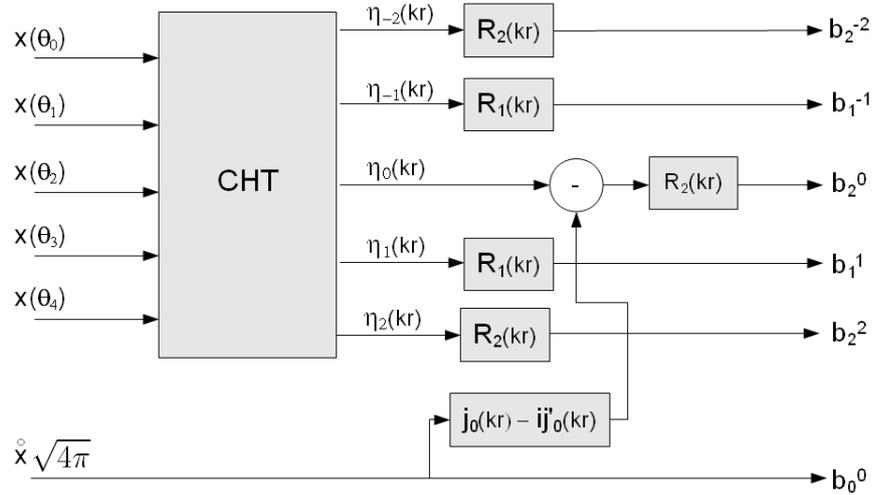


Figure 5.5: Block diagram of decomposition and separation process for a non-coincident microphone array consisting of 6 microphones .

Omnidirectional vs. cardioid microphones: The mentioned array design example uses cardioid microphones to sample the wave field on the circle. Of course it is also possible to use omnidirectional microphones instead. This approach is also described in [ME08b]. The main difference between the two approaches regards the different holography filters $R_n(kr)$

$$R_n(kr) = \begin{cases} \sim \frac{1}{j_n(kr)}, & \text{omnidirectional microphone} \\ \sim \frac{1}{j_n(kr) - ij'_n(kr)}, & \text{cardioid microphone} \end{cases}. \quad (5.22)$$

For omnidirectional elements an inversion of the Bessel function is required. This is not practicable for all frequencies because of singularities at high frequencies, see fig. 5.1. Further, the higher order filters exhibit a low shelving behaviour with a slope of n 20dB/decade.

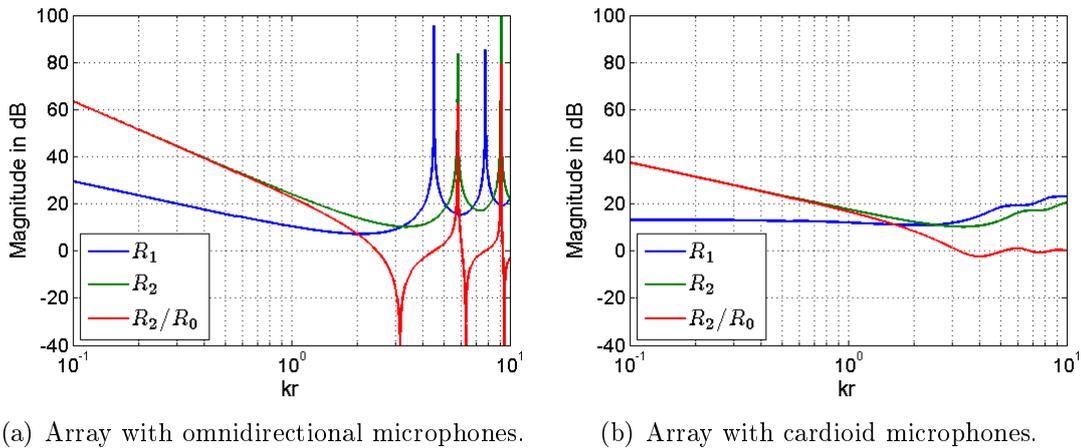


Figure 5.6: Radial compensation filters.

When comparing fig. 5.6(a) with fig. 5.6(b) it becomes obvious that cardioid microphones offer preferable characteristics with respect to the implementation of the radial compensation filters as the singularities at high frequencies are eliminated and the low shelving can be reduced to a slope of $(n - 1)$ 20dB/decade. But it is also well known that cardioid microphones are usually not as well matched as omnidirectional microphones which is important in array processing. And additionally, the orientation of the cardioid microphones needs to be precise in order to avoid a mismatch between the analytic model and the praxis, see section 6.

5.2.5 Higher order arrays

Higher order arrays can be designed by increasing the number of elements on the outer concentric ring and by increasing the total number of sub-rings.

Generally, the number of spherical harmonics coefficients that can be controlled by a N^{th} order array is given by

$$No = \frac{(N + 1)(N + 2)}{2}. \quad (5.23)$$

As rule of thumb it can be stated that the number of elements needed to correctly analyse the wave field, must be at least as high as the number of controllable spherical harmonics. Further, each concentric sub-ring of order n must consist of at least $2n + 1$ elements in order to avoid spatial aliasing.

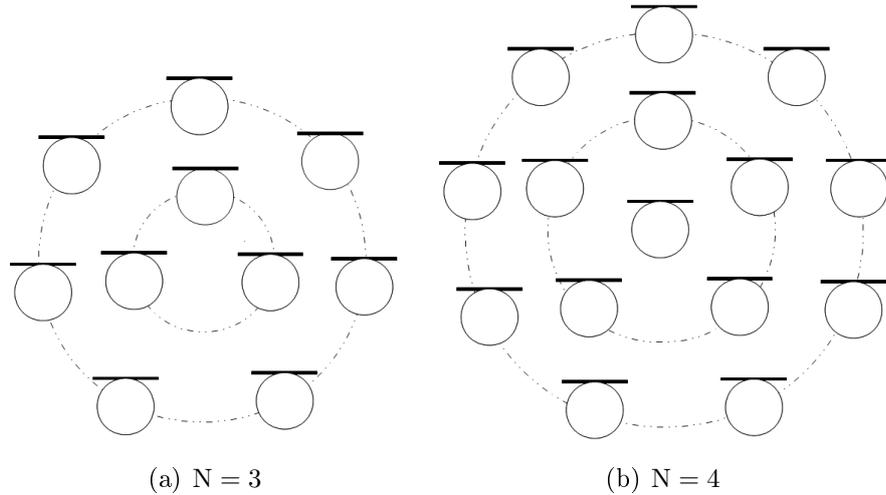


Figure 5.7: Higher order arrays.

Let us consider a third-order order array, see fig. 5.7(a). With (5.23) the number of controllable spherical harmonics is $No = 10$ and accordingly, at least 10 sampling elements are needed overall. The outer circle needs to be sampled with at least 7 elements. As can be seen in figure 5.2 the circular modes for $m = -1$, $m = 0$, $m = 1$ are aliased. For the separation one can use a first-order sub-ring with 3 elements. Note that this approach is only valid in a frequency band that is limited by the upper cut-off frequency f_u , which changes for each ring as the radius and the order changes (see eq. (5.9)).

A fourth-order array consists of at least 15 elements, whereas the outer circle is sampled at 9 points, see fig. 5.7(b). The circular modes for $m = -2$ to $m = 2$ are aliased in this example. Therefore, an inner circle with 5 microphones and a centre element is needed to separate the aliased mode for $m = 0$.

General design instructions for a N^{th} order circular array can be drawn by emphasizing and summarizing the previously mentioned and are also given in [ME08b]

- The total number of needed elements is given by $M \geq No$, see eq. (5.23)
- The total number of concentric rings is

$$= \begin{cases} \frac{N}{2} + \text{center element} , & \text{for even orders} \\ \frac{(N+1)}{2}, & \text{for odd orders} \end{cases} \quad (5.24)$$

- The number of elements on a n^{th} order sub-ring is $2n + 1$.

Note that the centre sensor needs to be an omnidirectional microphone.

6. Compare model with measurement

Ideally, the frequency independent encoder matrix should decompose the sampled sound field into separated but frequency dependent modes. As this is not the case, it is evident that the ideal analytic model is not consistent with the measured data. A link between the ideal model and the actual measured system is created on the basis of the MIMO-system in the SH domain. The definition of the MIMO-system is given by eq. (4.6), but as the further deliberations are based on this definition it is recalled here

$$\boldsymbol{\chi}_{N_A} = \mathbf{C}_{N_A}^+ \mathbf{H} \mathbf{D}_{N_C}^+ \boldsymbol{\phi}_{N_C}. \quad (6.1)$$

With eq (5.19) and eq. (A.32) the MIMO-system with the wave spectra on both sides is defined as

$$\mathbf{b}_{N_A} = (\mathbf{F}_{N_A} \circ \mathbf{C}_{N_A})^{-1} \mathbf{H} \mathbf{D}_{N_C}^+ \mathbf{S}_{N_C} \mathbf{b}_{N_C}, \quad (6.2)$$

where \mathbf{F}_{N_A} contains the frequency dependent propagation terms $\rho_n(kr_j)$, \circ denotes the Hadamard product of two matrices, $\mathbf{S}_{N_C} = \text{diag}_{N_C} \{(-ik h_n(kr_i))^{-1}\}$ and \mathbf{b} denotes the wave spectra in vector notation.

Note that the matrix \mathbf{F}_{N_A} for an microphone array with equal microphone types and equal distances to the centre (cf. all microphones are positioned on a concentric circle) is a simple diagonal matrix (see eq. (A.40)), but for the array prototype described in 3, which consist of microphones with different directivity patterns and a centre element, each row of the matrix is individually defined by

$$\text{vec}_{N_A} \{\rho_n(kr)\} = \left(\rho_0, \rho_1, \rho_1, \dots, \underbrace{\rho_n, \dots, \rho_n}_{n+1}, \dots, \rho_{N_A} \right). \quad (6.3)$$

In order to investigate the differences of the analytical model and the actual measured data eq. (6.2) is rearranged so that the model includes all analytic parameters and

the measurement model comprises excitation and the measured transmission paths

$$\underbrace{(\mathbf{F}_{N_A} \circ \mathbf{C}_{N_A}) \mathbf{S}_{N_A}^{-1}}_{\text{model}} \mathbf{b}_{N_A} = \underbrace{\mathbf{H} \mathbf{D}_{N_C}^+}_{\text{measurement}} \mathbf{b}_{N_C}. \quad (6.4)$$

The adaptability of the model is improved by adding adjustment parameters in the analytic model. In practice, the directivity patterns of microphones may deviate from the ideal patterns and also vary over frequency. In order to account for that effect the radial dependent propagation terms of the cardioid microphones are adjusted to

$$\tilde{\rho}_n(kr) = \frac{1}{2} [\alpha j_n(kr) - (2 - \alpha) i j_n'(kr)], \quad (6.5)$$

where $0 \leq \alpha \leq 2$ sets the proportion of the omnidirectional and figure-of-eight characteristics ($\alpha = 2$ results in omni directional characteristics).

Further, smallest deviations in position or radial orientation can cause noticeable differences in the results. Consequently, an angle offset parameter $\Delta\varphi_j$ is included in order to adjust the model. The last parameter g_{model} is a scalar used to compensate for arising gain mismatches between model and measured data.

Insertion of all mentioned parameter yields

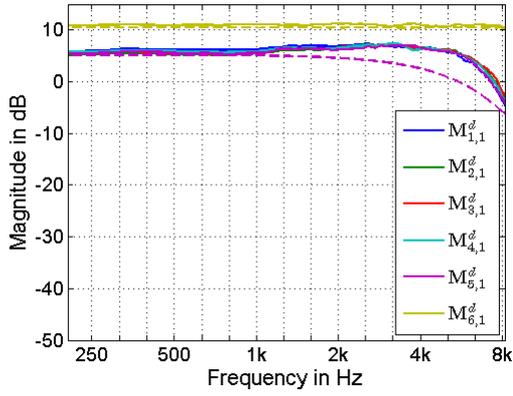
$$\left[g_{model} \tilde{\mathbf{F}}_{N_A} \circ \tilde{\mathbf{C}}_{N_A} \right] \mathbf{S}_{N_A}^{-1} \mathbf{b}_{N_A} = \mathbf{H} \mathbf{D}_{N_C}^+ \mathbf{b}_{N_C}, \quad (6.6)$$

where the tilde denotes the adjusted model matrices.

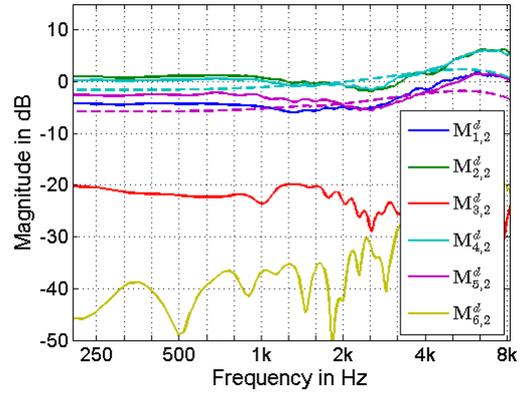
Figure 6.1 shows the column-wise comparison of the model and the measurement matrix in the array achievable subspace ($\mathbf{H} \mathbf{D}_{N_C}^+ = \mathbf{M}_{N_A}^d$). The slashed lines indicate model data and the full lines measurement data and if model lines are missing this is either because they are overlaid with others or they are below the lower display limit of $-50dB$. The best fitting model was empirically found for $\alpha = 1.1$ and $g_{model} = 20$. Note that a rough estimation of $\Delta\varphi_j$ is discussed later.

Generally, the measurement matrix can be interpreted as the matrix that links the excited spherical harmonics with the microphone signals. And the inverse of the model matrix describes the decomposition of the microphone signals into the wave spectra \mathbf{b}_{N_A} . Consequently the product of the inverse model and the measurement matrix should yield an identity matrix.

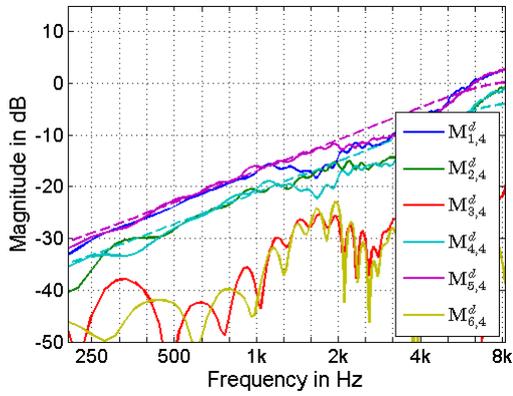
However, each column represents the expected (model) and the actually measured microphone signals due to an excitation of a specific spherical harmonic. Accordingly, fig 6.1(a) shows the microphone signals due to an excitation of the 'omnidirectional' spherical harmonic (b_0^0). It can be seen that the model approximately fits the measured data up to $2kHz$, whereas the measured data exceeds the model above that boundary. In fig. 6.1(b) it is illustrated that the signals of microphones 3 and 6 exceed the model which indicates that the directivity characteristics or the orientation of the microphones are not ideal. Essentially the same effect is depicted in fig. 6.1(c). However, the most significant mismatch is observed in 6.1(d) as an excitation of the upward facing spherical harmonic causes too high levels at all microphones. This observation suggests that the this harmonic is present in the directivity patterns of all microphones and consequently the microphones may obey an upward facing directivity. Further, this means that the measured decomposition and normalisation filters need less amplification when compared to the analytic filters (see. ch 4). The microphone directivity patterns are analysed in sec. 6.1.



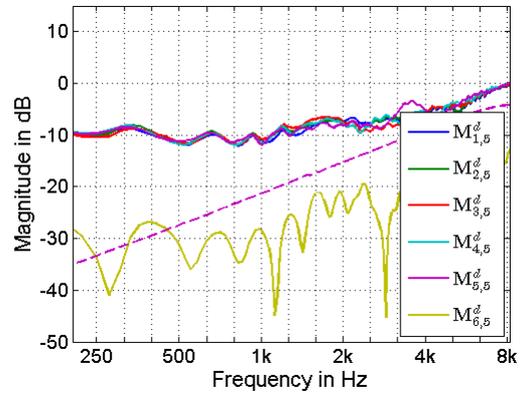
(a) First column of the model and measurement data.



(b) Second column. Magnitude of mic. 3 and 6 ideally $\rightarrow -\infty$.



(c) Fourth column. Magnitude of mic. 3 and 6 ideally $\rightarrow -\infty$.



(d) Fifth column. Magnitude of mic. 6 ideally $\rightarrow -\infty$.

Figure 6.1: Comparison of model and measurement data for $g_{model} = 20$ and $\alpha = 1.1$.

6.1 Analysing the microphone directivity patterns

As an initial step, the main axis of each cardioid microphone is rotated to an azimuth angle of $\varphi_j = 0^\circ$, which enables the comparability of the individual microphones. Further, it is necessary to ensure that the acoustic centre of each microphone coincides with the origin of the coordinate system under consideration because a dislocation, or a coaxial translation from the origin would lead to incorrect spherical harmonics spectra [Bau11].

The alignment of the acoustic centre to the coordinate origin is done based on the array geometry. Let us define the euclidean distance of the i^{th} loudspeaker to the j^{th} microphone as $\|(\mathbf{r}_j - \mathbf{r}_i)\|$ and the euclidean distance of i^{th} loudspeaker to the origin as $\|\mathbf{r}_i\|$ where $\|\cdot\|$ denotes the euclidean distance and \mathbf{r}_j and \mathbf{r}_i the position of microphone j and loudspeaker i in Cartesian coordinates. Thus, the distance offset is given by $\|(\mathbf{r}_j - \mathbf{r}_i)\| - \|\mathbf{r}_i\|$ and the realignment is done as described in 3.

The spherical harmonics spectrum of the microphone directivity patterns are obtained by applying the DSHT to the realigned measured impulse response matrix \mathbf{H} .

It can be seen in fig. 6.2(a) the directivity of the cardioid microphone ($j=1$) mainly consists of the 'omnidirectional' (Y_0^0) and the 'figure-of-eight' (Y_1^1) spherical harmonic. But, it is also evident that the ratio between those two components and consequently the overall directivity changes over frequency. Further, fig. 6.3 depicts differences

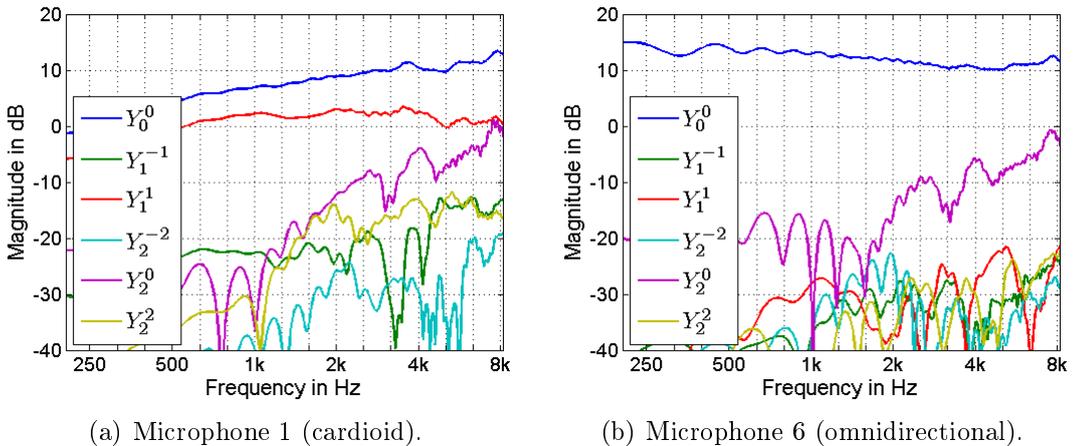


Figure 6.2: Spherical harmonics decomposition of button directivity characteristics.

between individual cardioid microphones which additionally complicates the definition of a universally valid analytic model.

Moreover, figures 6.2(a) and 6.2(b) confirm the supposed increased sensitivity towards sound impinging from vertical directions and the centre element is not ideally omnidirectional.

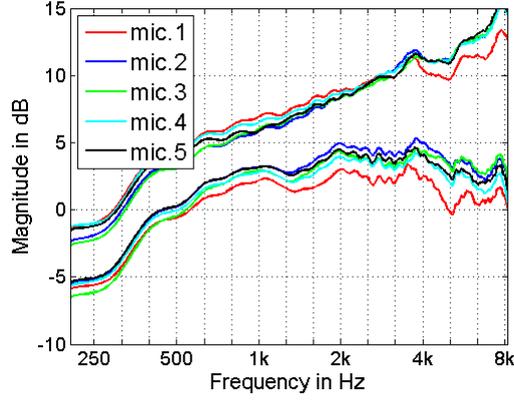


Figure 6.3: Omnidirectional (upper lines) and figure of eight spherical harmonic (lower lines) components of all cardioid microphones .

Estimation of positioning error: Generally, an cardioid directivity is obtained by a combination of an omni directional and a figure-of-eight characteristics. Accordingly, the rotated and aligned directivity patterns are ideally obtained by a superposition of the first and third spherical harmonic whilst other harmonics should not be present in the decomposition and thus the corresponding coefficients are ideally around zero.

A measure for estimating the angel offset between the ideal positions of the cardioid microphones and the actual positions is given by the ratio of the second harmonic and the third harmonic. As stated in eq. (5.4) the second harmonic corresponds to $\sin(\varphi)$ and the third harmonic to $\cos(\varphi)$. Therefore the theoretical angle offset in azimuth direction for idealized considerations is obtained by

$$\Delta\varphi_j = \arctan \left(\frac{\sin(\Delta\varphi_j)}{\cos(\Delta\varphi_j)} \right) \approx \left(\frac{|Y_1^{-1}|}{|Y_1^1|} \right)_j, \quad (6.7)$$

where $\Delta\varphi_j$ denotes the misalignment of the j^{th} microphone. (The approximation ¹ is valid as long as the third harmonic is at least three times higher than the second.) Since this estimation is only valid if the characteristics are assumed to be ideal the calculated angle offsets have to be interpreted with caution if the characteristics are not ideal.

6.2 Holographic error

In general, there are two potential error sources that can cause the mismatch between the analytic model and the actual measurements. One of them is related to the finite resolution sampling and the application of the DSHT A.5 with the well studied truncation error² and the spatial aliasing error³.

The other error sources are array imperfections such as microphone positions, gain mismatches between microphones and in our case also the frequency dependent change of microphone directivity.

An error measure that is able of evaluating possible aliasing errors as well as the mentioned array imperfections is presented in [Ple09] and is denoted as the holographic error $\boldsymbol{\sigma}$ and describes the error between an actual sound source and its biased replica from holography. The determination given in [Ple09] has to be adapted to be usable for the evaluation of the holographic error of a circular array with a centre element and is given by

$$\boldsymbol{\sigma}_{N_A, N_C} = [\mathbf{P}_{N_A}^{-1}(\mathbf{F}_{N_A} \circ \mathbf{C}_{N_A})^{-1} \mathbf{C}_{N_C} \circ \mathbf{F}_{N_C} \mathbf{P}_{N_C} - [\mathbf{I}_{N_A} \mathbf{0}]] \mathbf{y}_{N_C}(\boldsymbol{\theta}_i), \quad (6.8)$$

where $\mathbf{P}_N \mathbf{y}_N(\boldsymbol{\theta}_i) = \mathbf{b}_N$ is used to represent the wave spectrum of a source impinging from $\boldsymbol{\theta}_i$.

The normalized scalar holographic error $\|\boldsymbol{\sigma}_{N_A, N_C}\|$ is defined by

$$\|\boldsymbol{\sigma}_{N_A, N_C}\|^2 = \frac{1}{4\pi} \frac{1}{\|\mathbf{y}_{N_A}\|^2} \|\mathbf{P}_{N_A}^{-1}(\mathbf{F}_{N_A} \circ \mathbf{C}_{N_A})^{-1} \mathbf{C}_{N_C} \circ \mathbf{F}_{N_C} \mathbf{P}_{N_C} - [\mathbf{I}_{N_A} \mathbf{0}]\|_F^2, \quad (6.9)$$

¹ $\arctan\left(\frac{|Y_1^{-1}|}{|Y_1^1|}\right) \approx \left(\frac{|Y_1^{-1}|}{|Y_1^1|}\right)$

²Components of higher order are not included in the decomposed spectrum.

³Higher order components are mirrored into the lower order harmonics

where $\|\mathbf{y}_{N_A}\|^2$ denotes the squared norm of the spherical harmonics vector. Please note that the energy of the spherical harmonics vector is not constant over the zenith angle in the case where just the symmetrical harmonics are considered and therefore, the normalization and accordingly the scalar holographic error is dependent on the considered latitude.

Aliasing errors are included by setting $N_A < N_C$ gain mismatches are examined by including a diagonal matrix \mathbf{G} and changing of the arguments of the spherical harmonics matrix $\tilde{\mathbf{C}}_{N_A}$ simulates an inaccurate microphone placing, see $\Delta\varphi_j$. Further, the holographic error is applicable to analyse the effect of not ideal microphone characteristics by changing the parameter α in $\tilde{\mathbf{F}}_{N_A}$. Consequently, the holographic error converges towards zero if $N_A = N_C$ and $(\tilde{\mathbf{F}}_{N_A} \circ \tilde{\mathbf{C}}_{N_A})^{-1} = \mathbf{C}_{N_C} \circ \mathbf{F}_{N_C}$.

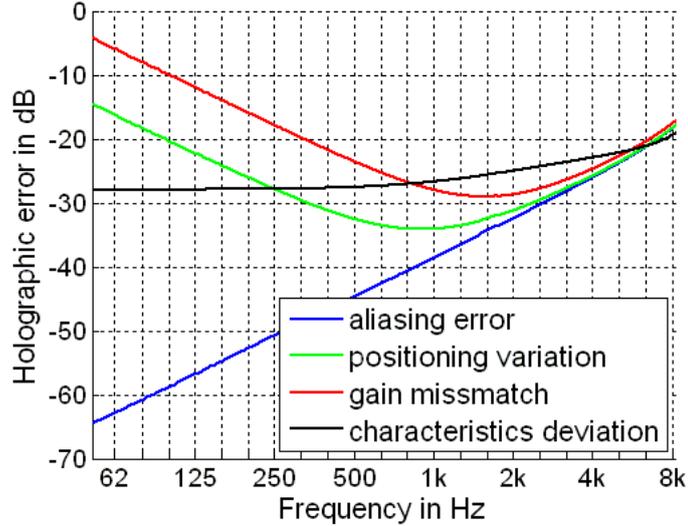


Figure 6.4: Holographic error for the microphone array presented in ch. 3 for $N_C = 13$ and normalized for $\vartheta = \pi/2$. Positions are varied randomly $-4^\circ \leq \Delta\varphi_j \leq 4^\circ$, gains are varied randomly $1 \leq \mathbf{G}_{i,i} \leq 1.2$ and $\alpha = 1.4$.

Figure 6.4 shows the influence of the aforementioned simulated errors. It can be clearly seen that the aliasing error increases with frequency and that it dominates in the frequency range above $6kHz$. In contrast, the other error sources have higher impact on the low frequency performance, which is caused by an amplification of any error with the high gains of the inverse holography filters in the lower frequency range; see fig. 5.6.

7. Summary and future work

This thesis dealt with modal beamforming using circular microphone arrays. After deriving certain pick-up pattern optimization weightings and summarizing the concept of modal beamforming in chapter 2, the main focus has been on the decomposition of the sound-field into orthogonal and directive modes in the spherical harmonics domain that are the targeted signals of the modal beamformer.

An approach that is based on measurement data has been presented in chapter 4. Even though, this concept requires extensive measurements it allows for the implementation of a working beamforming-system without extensive expertise in the field of acoustic holography or sound-field theory. The practical implementation of the system using a 2nd order array prototype has been also discussed in chapter 4. It has been shown that a regularized system-matrix inversion using the SVD is necessary to improve the robustness and consequently, the applicability of the presented beamforming-system. Further, it has been demonstrated that the initial measurement complexity can be reduced without significantly degrading the overall performance.

The analytic description of circular array based modal beamforming has been discussed in chapter 5. This chapter mainly dealt with the separation of aliased modes and the effect of different microphone types on the radial compensation filters (holography filters). It has been shown that circular arrays consisting of cardioid microphones obey preferable characteristics concerning the implementation of the aforementioned compensation filters. Further, general design rules in order to design higher order arrays has been given.

In chapter 6 it has been shown that there are noticeable differences between the analytical model and the practical measured data. In order to explain the occurring

mismatches, potential error sources such as array imperfections has been analysed. Based on the observations, the main reason for the mismatch is assumed to be due to the non ideal directivity characteristics of the used array microphones.

7.1 Future work

In order to further improve the robustness of the beamforming-system the weighting coefficients could be designed for a constant WNG under the constraint of maximizing the DI. However, it is not assumed that this would significantly increase the performance as the filter regularization has a similar effect.

The overall implemented system consists of 36 filters that decompose the microphone signals into directive modes. Further research may deal with the reduction of the needed filters. An obvious solution would be to omit filters corresponding to IRs with insignificant amplitudes.

As pointed out in chapter 6 array imperfections can degrade the performance. Consequently, a microphone array consisting of high-quality sensors that are extremely accurate placed could allow for implementing the beamforming-system based on the analytic model. Thus, the measurements of each individual array would become needless. However, it has been shown in chapter 4 that measurements from 6 directions could be sufficient.

Further, the implementation of higher order arrays is considered to be subject to future research.

A. Sound field description in spherical coordinates

This chapter provides an overview of the theory of sound field representation in the spherical coordinate system. Further, the spatial Fourier transform and the characteristic of spherical harmonics are discussed. Throughout this entire text the spherical coordinate system is used instead of the more common Cartesian system to describe the geometrical and spatial relations. The transformation of the Cartesian (x, y, z) to the spherical coordinate system (r, φ, ϑ) is defined by

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = r \begin{pmatrix} \cos(\varphi) \sin(\vartheta) \\ \sin(\varphi) \sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix} = r\boldsymbol{\theta} \quad (\text{A.1})$$

and the graphical interpretation is provided by figure A.1. The corresponding angles are the inclination or zenith angle limited between $0 \leq \vartheta \leq \pi$ and the azimuthal angle limited between $0 \leq \varphi \leq 2\pi$. The following context, relations and mathematical framework is mainly build upon [Wil99], [BSG⁺89] and [Zot09].

A.1 Sound field theory

As is known, the two most prominent equations to describe the acoustic properties of a sound field in the time-domain are the linear lossless wave equation for the pressure $p(\mathbf{r}, t)$ and the Euler equation for the particle velocity $\mathbf{v}(\mathbf{r}, t)$. According to [Wil99]

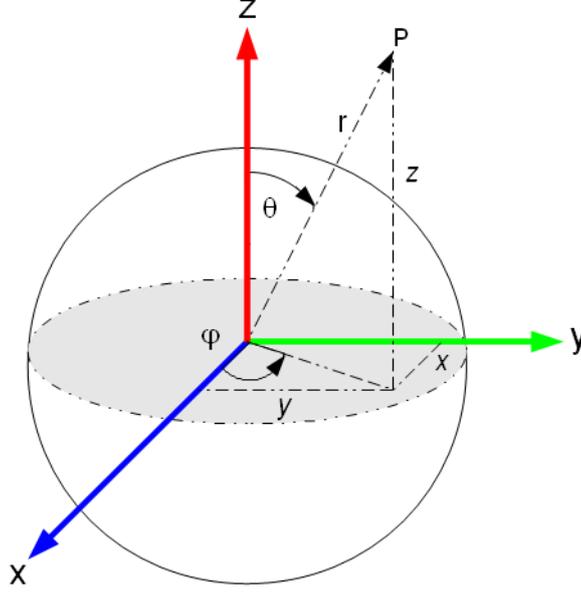


Figure A.1: Relation between Cartesian and spherical coordinate system.

these equations for a homogeneous fluid with no viscosity can be written as

$$\Delta p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0 \quad (\text{A.2})$$

$$\varrho_0 \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial t} = -\nabla p(\mathbf{r}, t), \quad (\text{A.3})$$

where \mathbf{r} and t are the spatial location and time of the field point, Δ is the Laplacian, c is the speed of sound in the medium and ϱ_0 is the air density.

In the frequency domain, the wave equation simplifies to the well known Helmholtz equation and equations (A.2) and (A.3) can be rewritten to

$$(\Delta + k^2)p(\mathbf{r}, \omega) = 0 \quad (\text{A.4})$$

$$\frac{\varrho_0 c}{i} k \mathbf{v}(\mathbf{r}, \omega) = -\nabla p(\mathbf{r}, \omega), \quad (\text{A.5})$$

where Δ again is the Laplacian, which is for the spherical coordinate system and a function f defined by

$$\begin{aligned} \Delta f &= \Delta_r f + \Delta_\varphi f + \Delta_\vartheta f \\ &= \frac{1}{r^2} \frac{\partial^2 r^2}{\partial r^2} + \frac{1}{r^2 \sin^2(\vartheta)} \frac{\partial^2 f}{\partial \varphi^2} + \frac{1}{r^2 \sin(\vartheta)} \frac{\partial}{\partial \vartheta} (\sin(\vartheta) \frac{\partial f}{\partial \vartheta}). \end{aligned} \quad (\text{A.6})$$

When using a product ansatz and a separation of the variables (eq. (A.8))

$$p(\mathbf{r}) = R(kr)\Phi(\varphi)\Theta(\vartheta) \quad (\text{A.7})$$

it is possible to simplify the homogeneous Helmholtz equation into three decoupled differential terms which are known from literature [BSG⁺89] and [Zot09]

$$\begin{pmatrix} \frac{\partial^2}{\partial(kr)^2}(kr)^2 R(kr) + [(kr)^2 - n(n+1)] R(kr) \\ \frac{\partial^2}{\partial\varphi^2}\Phi(\varphi) + m^2\Phi(\varphi) \\ \frac{\partial}{\partial\mu} \left[\frac{1}{1-\mu^2} \frac{\partial}{\partial\mu} \Theta(\vartheta) \right] + \left[n(n+1) - \frac{m^2}{1-\mu^2} \right] \Theta(\vartheta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (\text{A.8})$$

where $\mu = \cos(\vartheta)$. The obtained independent differential equations are a spherical Bessel differential equation for the radial term, a linear differential equation for the azimuth term and an associated Legendre differential equation for the zenith dependence of the sound field. Section A.2 discusses the solutions of the differential equations briefly.

A.2 Solution of the Helmholtz equation

As mentioned above the homogeneous Helmholtz equation describes just source free sound fields. Therefore, the source free fields and the source areas are separated by a spherical boundary on which the acoustic field quantities are defined or measured. According to the Kirchhoff-Helmholtz integral and by using the angular and radial solutions it is possible to describe the source free part by knowing the values on the boundary. The boundary value problem can then be split into three main types [Zot09]

- Exterior problem: The boundary is enclosing all sources and the aim is to determine the outer field, hence exterior problem.
- Interior problem: The aim is to determine the sound field that is caused by sources outside of the boundary.
- Mixed problem: The previously mentioned conditions can be combined into a mixed problem. This means that there are sources inside and outside a source free volume or a source free field is inside and outside of a source volume.

Anyway, the solution of such problems relies on two different boundaries where the field quantities are known.

In the following, just solutions of eq. (A.8) that guarantee a description of the source free field without singularities are selected.

Radial solution: Possible solutions to the radial equation would be the spherical Bessel, Neumann and Hankel functions of the first and second kind

$$R(kr) = R_n(kr) = \left\{ \begin{array}{l} j_n(kr), y_n(kr) \text{ for interior boundary problems} \\ h_n^{(1)}(kr), h_n^{(2)}(kr) \text{ for exterior boundary problems} \end{array} \right\}, \quad (\text{A.9})$$

where n is referred to as the order. For interior boundary problems the radial function must be regular inside the bounded area. The only feasible solution that meets this requirement is the spherical Bessel function, because it is regular for all $kr \geq 0$ (figure A.2 depicts the behaviour of both functions and illustrates the singularity of the Neumann function for small kr). For exterior problems the spherical Hankel function

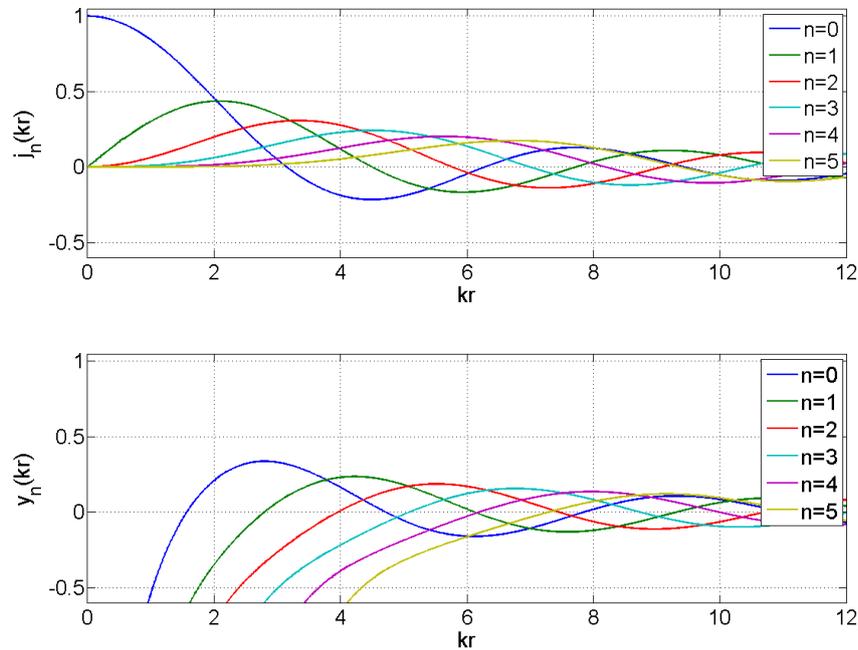


Figure A.2: The spherical Bessel and Neumann functions up to order $n = 5$.

of the second kind is chosen because it yet alone satisfies the Sommerfeld radiation

condition. These Hankel functions are a complex valued composite of the previously mentioned functions $j_n(kr)$ and $y_n(kr)$

$$\begin{aligned} h_n^{(1)}(kr) &= j_n(kr) + iy_n(kr) \\ h_n^{(2)}(kr) &= h_n^{(1)*}(kr). \end{aligned} \quad (\text{A.10})$$

When taking a closer look on figure A.3 (magnitude of Hankel functions) it can be seen that for small kr higher order components decay rapidly with increasing kr and r respectively (near-field region) and that from a specific kr all orders share the same amount of radial decay (far-field region).

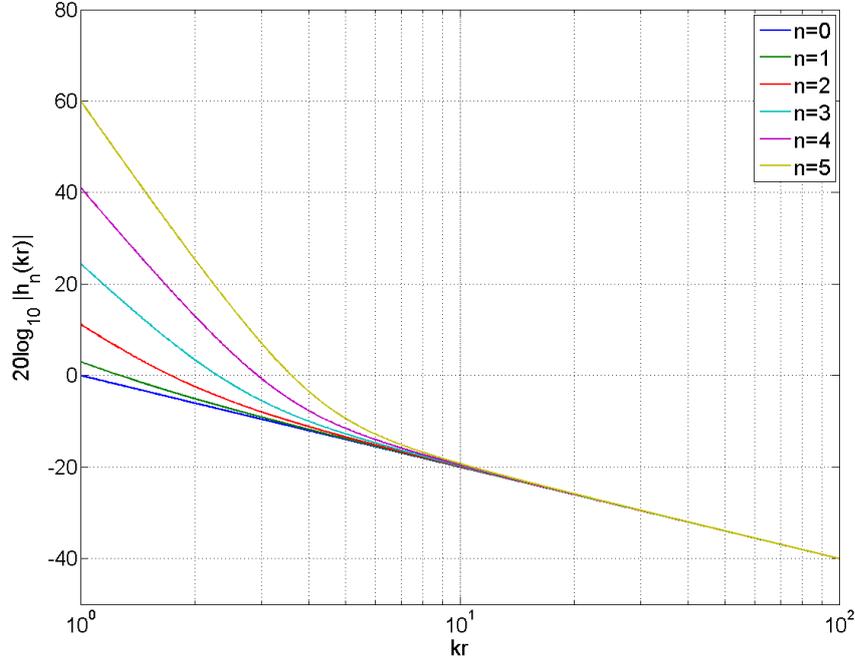


Figure A.3: The magnitude of the spherical Hankel functions up to order $n = 5$.

Zenith solution: The Legendre differential function can be solved by the Legendre functions of the first and second kind.

$$\Theta_n^m(\vartheta) = \left\{ \underline{P}_n^m(\mu), Q_n^m(\mu) \right\}, \quad (\text{A.11})$$

where m denotes the degree. Again the regularity of the function over the region of interest is crucial. Accordingly, the Legendre functions of the second kind are omitted due to their singularity at $\mu = \pm 1$, what equals $\vartheta = 0^\circ$ and 180° .

Azimuth solution: The linear differential equation is solved by sines and cosines or complex exponentials

$$\Phi_m(\varphi) = \begin{cases} \sin(m\varphi), \cos(m\varphi) \text{ real valued} \\ e^{\pm im\varphi} \text{ complex} \end{cases}, \quad (\text{A.12})$$

whereas the selection does not affect the applicability of the angular solution.

Combination of the angular solutions: For the sake of convenience, both angular solutions are combined to form the spherical harmonics which depend on both angles φ and ϑ and are denoted per definition by

$$Y_n^m(\boldsymbol{\theta}) = N_n^{|m|} \Phi_m(\varphi) P_n^{|m|}(\vartheta), \quad (\text{A.13})$$

where $N_n^{|m|}$ is a scalar orthonormalization constant. With equations (A.11) and (A.12) it is possible to rewrite eq. (A.13) and the real or complex valued spherical harmonics to

$$Y_n^m(\boldsymbol{\theta}) = N_n^{|m|} P_n^{|m|}(\cos(\vartheta)) \begin{cases} \sin(m\varphi), & \text{for } m < 0 \\ \cos(m\varphi), & \text{for } m > 0 \\ e^{\pm im\varphi}. \end{cases} \quad (\text{A.14})$$

Total solution: According to [Wil99], [GD04] and [Gir96] the total solution of the Helmholtz equation is then defined by

$$p(kr, \boldsymbol{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [b_n^m j_n(kr) + c_n^m h_n^{(2)}(kr)] Y_n^m(\boldsymbol{\theta}), \quad (\text{A.15})$$

with b_n^m and c_n^m as the wave spectra of the incident and the radiating field respectively and $Y_n^m(\boldsymbol{\theta})$ as the normalized spherical harmonics.

A.3 Spherical harmonics

Just as the Fourier basis represents an important tool for the evaluation of a one- or two-dimensional function, the spherical harmonic basis is a similar tool but defined

on the surface of a sphere. The real or complex valued functions are defined by eq. (A.14) where the indices $n \in \mathbb{N}_0$ and $m \in \mathbb{Z} : -n \leq m \leq n$ are again labelled as *order* and *degree*. In order to serve as a basis of any spherical function, the spherical harmonics must be orthonormal. This orthonormality condition can be written using a product of Kronecker deltas:

$$\int_{S^2} Y_{n'}^{m'}(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} = \delta_{nn'} \delta_{mm'}, \quad (\text{A.16})$$

where δ_{ij} denotes the Kronecker's delta. The integral $\int_{S^2} d\boldsymbol{\theta} = \int_0^{2\pi} \int_0^\pi \sin \vartheta d\vartheta d\varphi$ covers the entire surface of the unit sphere, denoted by S^2 . In words, eq. (A.16) states that the inner product of two spherical harmonics with different indices vanishes over the sphere and equals unity otherwise. With the normalization factor, which can be derived from (A.16) and is defined as

$$N_n^{m=} = \sqrt{\frac{(2n+1)(2-\delta_m)(n-m)!}{4\pi(n+m)!}}. \quad (\text{A.17})$$

the spherical harmonics are fully described and depicted in figure A.4. Summarized

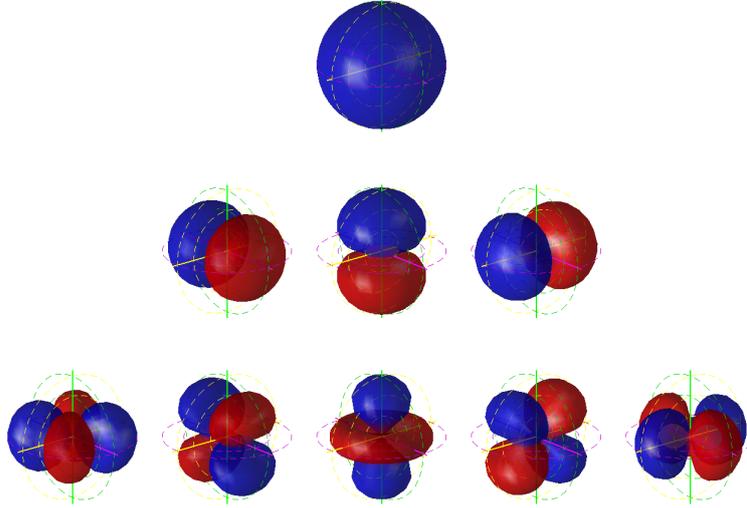


Figure A.4: Spherical harmonic functions up to order $n=3$ [PZ08].

spherical harmonics describe the orthonormal basis of the angular part of the spherical coordinate system and can be seen as spherical eigenfunctions or modes.

A.4 Spherical harmonics transform

In general, orthonormal basis functions allow the expression of any piecewise continuous function over a certain interval as a linear combination of the infinite series of linearly independent base functions. Accordingly, the basis functions can be seen as small fragments of information. By scaling and combining them it is possible to produce either the exact original function (for an infinite series of base functions or a band-limited spherical function) or the band-limited approximation of the source function (just a finite number of base functions is used while the source function also consist of higher frequencies or base functions of higher order).

In the spherical domain it is possible to transform any square integrable function $f(\boldsymbol{\theta})$ on the sphere into the spherical harmonics domain using the so called spherical harmonics transform

$$SHT\{f(\boldsymbol{\theta})\} = \int_{S^2} f(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} := \gamma_n^m. \quad (\text{A.18})$$

In words eq. (A.18) states that the spherical harmonics coefficients γ_n^m are defined by the correlation of the function $f(\boldsymbol{\theta})$ with the corresponding transform kernel, the respective spherical harmonic Y_n^m , over all possible angular positions. Further, it is possible to draw a comparison to the Fourier transform with the transformation kernel $e^{\pm i\omega t}$. Therefore, the spherical harmonic transform is often referred to as a spatial Fourier transform, where the indices n, m instead of the frequency variable ω are used to choose coefficients of the resulting spherical spectrum (totality of all γ_n^m). The inverse spherical harmonics transform (a full expansion without truncation) is defined by

$$ISHT\{\gamma_n^m\} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \gamma_n^m Y_n^m = f(\boldsymbol{\theta}) \quad (\text{A.19})$$

is complete

$$ISHT\{SHT\{f(\boldsymbol{\theta})\}\} = f(\boldsymbol{\theta}) \quad (\text{A.20})$$

and fulfils the Parseval theorem

$$\int_{S^2} |f(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} = \sum_{n=0}^{\infty} \sum_{m=-n}^n |\gamma_n^m|^2 \quad (\text{A.21})$$

that states that the transform is unitary (the integral of the square of a function is equal to the sum of the square of its transform).

Spherical wave spectra and wave spectra: The application of the spherical harmonics transform to the acoustic field quantities results in the spherical wave spectra [Wil99]

$$\psi_n^m(kr) = SHT\{p(kr, \boldsymbol{\theta})\} \quad (\text{A.22})$$

for the pressure and

$$\nu_n^m(kr) = SHT\{v_r(kr, \boldsymbol{\theta})\} \quad (\text{A.23})$$

for the radial component of the sound particle velocity at a specific radius r . The expansion of the spherical wave spectra is done according to eq. (A.19)

$$p(kr, \boldsymbol{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \psi_n^m(kr) Y_n^m(\boldsymbol{\theta}). \quad (\text{A.24})$$

If one compares equations (A.15) and (A.24), the relation between the spherical wave spectrum ψ_n^m and the wave spectrum b_n^m , c_n^m becomes obvious

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n [b_n^m j_n(kr) + c_n^m h_n^{(2)}(kr)] Y_n^m(\boldsymbol{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \psi_n^m(kr) Y_n^m(\boldsymbol{\theta}). \quad (\text{A.25})$$

According to eq. (A.25) and the relation between the pressure and velocity defined by the Euler equation (A.3), the pressure and particle velocity spectra are defined by [Zot09]

$$\psi_n^m(kr) = b_n^m j_n(kr) + c_n^m h_n^{(2)}(kr) \quad (\text{A.26})$$

$$\nu_n^m(kr) = \frac{i}{\rho_0 c} \left[b_n^m j_n'(kr) + c_n^m h_n^{(2)'}(kr) \right]. \quad (\text{A.27})$$

If an interior problem is considered ($c_n^m = 0$), the spherical sound pressure distribution or particle velocity distribution due to sources outside of the spherical boundary (microphone radius) r is known or has to be measured. Under the assumption of a free incident field it is possible to calculate the wave spectra coefficients b_n^m of the interior field by

$$\begin{aligned} \psi_n^m(kr) &= b_n^m j_n(kr) \\ b_n^m &= \frac{\psi_n^m(kr)}{j_n(kr)}. \end{aligned} \quad (\text{A.28})$$

Since the spherical Bessel functions $j_n(kr)$ (figure A.2) exhibit zeros for specific frequencies or radii it is not possible to produce all orders n for those specific conditions. However, these divisions by zero must be avoided in practice.

Spherical source distribution: According to [Zot09] a continuous distribution of sources guarantees an excitation of the interior field without zeros. Thereby, the continuous source strength distribution function $f(\boldsymbol{\theta})$ excites the inhomogeneous Helmholtz equation at a radius r_e

$$(\Delta + k^2)p = -\frac{\delta(r - r_e)}{r^2}f(\boldsymbol{\theta}) \quad (\text{A.29})$$

The complete solution of this inhomogeneous problem can be written as

$$p(kr, \boldsymbol{\theta} | \phi_n^m) = -ik \sum_{n=0}^{\infty} \sum_{m=-n}^n \phi_n^m Y_n^m(\boldsymbol{\theta}) \begin{cases} h_n^{(2)}(kr_e) j_n(kr), & \text{for } r \leq r_e \\ j_n(kr_e) h_n^{(2)}(kr), & \text{for } r \geq r_e \end{cases}, \quad (\text{A.30})$$

where r_e is the excitation radius and ϕ_n^m is referred to as modal source strength, which is defined by

$$\phi_n^m = \int_{S^2} f(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (\text{A.31})$$

or if b_n^m can be measured it is possible to calculate the matching source distribution at r_e by

$$\phi_n^m |_{r_e} = \frac{i}{k h_n^{(2)}(kr_e)} b_n^m. \quad (\text{A.32})$$

A.5 Discrete spherical harmonics transform

Again it is possible to draw the comparison to the Fourier transform, where the discretized form (DFT) is used to analyze or interpolate discrete sampled data. Accordingly, the discrete spherical harmonics transform (DSHT) declares the calculation of the spherical wave spectra using a finite set of spherically sampled data. As previously mentioned an infinite number of spherical harmonics would be theoretically needed to calculate the exact transformation of a non band-limited function. However, this is not practical and therefore the original data has to be angularly band-limited to the maximum order N of the DSHT. Further, in practice the sound field quantities are

spatially sampled at the microphone positions $\theta_{\mathbf{a}}$ with the sample index $j = 1, \dots, M$ and M as the number of microphones. Thereby, the used angular sampling scheme must be suitable to calculate the DSHT. In other words it is required that the discrete orthonormality condition [Raf05]

$$\sum_{j=1}^M \alpha_a Y_n^{m'}(\theta_{\mathbf{a}}) Y_n^m(\theta_{\mathbf{a}}) = \delta_{n-n'} \delta_{m-m'}, \quad (\text{A.33})$$

where α_a is a real value which depends on the sampling scheme, holds for the chosen microphone positions. With N as the maximum order and M as the number of microphones, the actual DSHT of the sampled signal $f(kr, \theta_{\mathbf{a}})$ and its inverse transform (DISHT) are given by

$$\gamma_n^m = \sum_{j=1}^M \alpha_a f(kr, \theta_{\mathbf{a}}) Y_n^m(\theta_{\mathbf{a}}) \quad (\text{A.34})$$

$$f(kr, \theta_{\mathbf{a}}) = \sum_{n=0}^N \sum_{m=-n}^n \gamma_n^m Y_n^m(\theta_{\mathbf{a}}). \quad (\text{A.35})$$

A.5.1 Matrix and vector notation

In order to simplify the latter formulations and the subsequent descriptions it is suitable to introduce the vector/matrix notation conventions introduced by [Zot09]. Thereby, the summations are represented by matrix multiplications (compare to equations (A.25) and (A.35))

$$\begin{aligned} \left[\mathbf{b}_N \mathbf{J}_N(kr) + \mathbf{c}_N \mathbf{H}_N^{(2)}(kr) \right] \mathbf{Y}_N &= \boldsymbol{\psi}_N(kr) \mathbf{Y}_N \\ \mathbf{f} &= \boldsymbol{\gamma}_N(kr) \mathbf{Y}_N. \end{aligned} \quad (\text{A.36})$$

Spherical harmonics coefficients: The wave spectra coefficients $(b_n^m | c_n^m)$ and the spherical wave spectra coefficients γ_n^m or $(\psi_n^m | \nu_n^m)$ for the orders $0 \leq n \leq N$ are written as column vectors with $(N+1)^2$ elements and are ordered as

follows

$$\mathbf{b}_N = \text{vec}_N \{b_n^m\} =: \left(\begin{array}{c} b_0^0 \\ b_1^{-1} \\ \vdots \\ b_n^{-n} \\ \vdots \\ b_n^n \\ \vdots \\ b_N^N \end{array} \right) \quad (A.37)$$

Spherical Bessel and Hankel functions: Again it is possible to define vectors of the spherical Bessel and Hankel functions

$$\mathbf{j}_N = \text{vec}_N \{j_n\} = \left(j_0, j_1, j_1, j_1, \dots, \underbrace{j_n, \dots, j_n}_{2n+1}, \dots, j_N \right)^T \quad (A.38)$$

$$\mathbf{h}_N = \text{vec}_N \{h_n\} = \left(h_0, h_1, h_1, h_1, \dots, \underbrace{h_n, \dots, h_n}_{2n+1}, \dots, h_N \right)^T \quad (A.39)$$

Note that both functions are independent of the degree m . Therefore, the values for one specific order n are repeated $2n+1$ times. In order to solve eq.(A.36) the matrices

$\mathbf{J}_N(kr)$ and $\mathbf{H}_N(kr)$ are defined by

$$\mathbf{J}_N(kr) = \text{diag}_N \{j_n\} = \text{diag} \{\mathbf{j}_N\} = \begin{pmatrix} j_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & j_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & j_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & j_1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & j_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & j_N \end{pmatrix} \quad (\text{A.40})$$

$$\mathbf{H}_N(kr) = \text{diag}_N \{h_n\} = \text{diag} \{\mathbf{h}_N\} = \begin{pmatrix} h_0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & h_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & h_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & h_1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & h_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & h_N \end{pmatrix} \quad (\text{A.41})$$

Spherical harmonics: The spherical harmonics are evaluated for an angular position $\boldsymbol{\theta}$ up to order N and can be written into a vector

$$\mathbf{y}_N(\boldsymbol{\theta}) = \text{vec}_N \{Y_n^m(\boldsymbol{\theta})\} =: \begin{pmatrix} Y_0^0(\boldsymbol{\theta}) \\ Y_1^{-1}(\boldsymbol{\theta}) \\ \vdots \\ \left. \begin{matrix} Y_n^{-n}(\boldsymbol{\theta}) \\ \vdots \\ Y_n^n(\boldsymbol{\theta}) \end{matrix} \right\} 2n+1 \\ \vdots \\ Y_N^N(\boldsymbol{\theta}) \end{pmatrix}. \quad (\text{A.42})$$

Furthermore, in order to obtain the full spherical harmonics matrix \mathbf{Y}_N the vectors

are combined according to their sampling node indices j

$$\mathbf{Y}_N = \begin{bmatrix} \mathbf{y}_N^T(\boldsymbol{\theta}_1) \\ \mathbf{y}_N^T(\boldsymbol{\theta}_2) \\ \mathbf{y}_N^T(\boldsymbol{\theta}_3) \\ \vdots \\ \mathbf{y}_N^T(\boldsymbol{\theta}_M) \end{bmatrix}. \quad (\text{A.43})$$

Discrete spherical patterns: The samples taken at different angular positions are conveniently written as vector whereby the entries are again ordered according to the sample node index j

$$\mathbf{f} = \begin{bmatrix} f(\boldsymbol{\theta}_1) \\ f(\boldsymbol{\theta}_2) \\ \cdot \\ \cdot \\ f(\boldsymbol{\theta}_M) \end{bmatrix}. \quad (\text{A.44})$$

DSHT and DISHT in matrix notation: By utilizing the latter notations it is possible to rewrite equations (A.34) and (A.35) to

$$\boldsymbol{\gamma}_N = \mathbf{Y}_N^+ \mathbf{f} \quad (\text{A.45})$$

$$\mathbf{f} = \mathbf{Y}_N \boldsymbol{\gamma}_N, \quad (\text{A.46})$$

where $^+$ denotes the pseudo inverse.

A.5.2 Spatial aliasing and sampling schemes

As can be seen in eq. (A.46) the decomposition of the spherical function $f(\boldsymbol{\theta})$ into the spherical harmonics domain requires an inversion of the matrix \mathbf{Y}_N . According to [PZ08] the solvability of the equation highly depends on the form of \mathbf{Y}_N and thus on the number and distribution of the sampling points

- If $M = (N + 1)^2$ and the inverse \mathbf{Y}_N^{-1} exists: the system of equations can be solved exactly.
- If $M < (N + 1)^2$ and the inverse $(\mathbf{Y}_N^T \mathbf{Y}_N)^{-1}$ exists: system is under-determined and there are infinite possible exact solutions.
- If $M > (N + 1)^2$ and the inverse $(\mathbf{Y}_N^T \mathbf{Y}_N)^{-1}$ exists: system is overdetermined and there is accordingly no exact solution, but an approximated solution may be calculated.

Spatial Aliasing: As well known from the sampling of time-signals, the highest correct representable frequency depends on the used sampling frequency. This relation is well-defined by the Nyquist-Shannon sampling theorem that states that a band-limited sampled signal can be perfectly reconstructed from an infinite sample sequence if the sampling frequency exceeds twice the highest frequency present in the original signal. In order to reduce spatial aliasing two main conditions according to [PZ08], [Zot09] or [Zot08] must be satisfied

- the number of controllable spherical harmonics must not exceed the number of sampling points:

$$M \geq \begin{cases} (N + 1)^2 & \text{full set of basis functions} \\ \frac{(N+1)(N+2)}{2} & \text{just symmetric basis functions} \end{cases} \quad (\text{A.47})$$

- the sampling positions must be well distributed or the problem has to be well conditioned:

$$\kappa(\mathbf{Y}_N^T \mathbf{Y}_N) \ll \infty, \quad (\text{A.48})$$

whereby \mathbf{Y}_N is defined by eq. (A.43).

Another criterion to reduce distortions by spatial aliasing is mainly based on the aperture radius r of the sampling configuration and is introduced by [ZN07]. An upper cut-off frequency is given by

$$f_u \approx \frac{c}{2r}. \quad (\text{A.49})$$

In practice, the non band-limited functions are sampled with a finite number of samples, what can cause that different continuous functions become indistinguishable due

to identical sampling values. Especially, for narrow spherical functions (high harmonics) it occurs that aliased copies are mirrored into the lower harmonics. Please note that the conditioning of $\kappa(\mathbf{Y}_N^T \mathbf{Y}_N)$ as well as the number of needed sampling points M highly depends on the chosen spherical distribution of the sampling points.

Sampling the sphere: In [Raf05] or [PZ08] described possible sampling schemes would be an equiangle sampling, a Gaussian sampling, a nearly uniform sampling or an equidistant sampling (according to Platonic solids). As previously mentioned all these sampling schemes require band-limited bandwidth (harmonic order) in order to avoid aliasing. Actually, the different schemes offer a trade-off between the number of needed microphones to resolve a specific maximum order N and the simplicity of their arrangement.

B. Calculation of order weights

B.1 Maximizing the DI

Starting from eq. 2.20 the denominator $\int_{-1}^1 d(\mu)^2 d\mu$ needs to be minimized under the constraint of $d(1) = 1 \rightarrow \sum_n \hat{w}_n = 1$.

With eq. (2.15) and the orthogonality of the Legendre polynomials the denominator is rewritten to

$$\int_{-1}^1 d(\mu)^2 d\mu = \sum_{n=0}^N \hat{w}_n^2 \frac{2}{2n+1} \quad (\text{B.1})$$

and minimized using a Lagrange multiplier.

The Lagrange function for this optimization problem is written as

$$\Lambda(\hat{w}_n, \lambda) = \sum_n \frac{2\hat{w}_n^2}{2n+1} - \lambda \left(1 - \sum_n \hat{w}_n \right),$$

where λ denotes the Lagrange multiplier.

In a next step the derivatives are set to zero

$$\frac{\partial \Lambda(\hat{w}_n, \lambda)}{\partial \hat{w}_n} = \frac{4\hat{w}_n}{2n+1} + \lambda = 0 \quad (\text{B.2})$$

$$\frac{\partial \Lambda(\hat{w}_n, \lambda)}{\partial \lambda} = -1 + \sum_n \hat{w}_n = 0. \quad (\text{B.3})$$

Equation (B.2) implies $\hat{w}_n^{(opt)} = -\frac{\lambda}{4}(2n+1)$ and if substituted in eq. (B.3) the Lagrange multiplier is found to be $\lambda = -\frac{4}{(N+1)^2}$. Consequently, the weights $\hat{w}_n^{(opt)}$ that

maximize the DI under the $\sum_n \hat{w}_n = 1$ constraint are found to be

$$\hat{w}_n^{(opt)} = \frac{2n+1}{(N+1)^2}. \quad (\text{B.4})$$

B.2 Maximizing front-to-back centroid

By substituting eq. (2.15), eq. (B.1) and the recursion formula of Legendre polynomials into eq. (2.22) the front-to-back centroid of a pick-up pattern is expressed as

$$r_E = \frac{\int_{-1}^1 \sum_{n'} \hat{w}_{n'} \left(\frac{n'+1}{2n'+1} P_{n'+1}(\mu) + \frac{n'}{2n'+1} P_{n'-1}(\mu) \right) \sum_n \hat{w}_n P_n(\mu) d\mu}{\sum_n \hat{w}_n^2 \frac{2}{2n+1}}. \quad (\text{B.5})$$

By emphasizing the orthogonality of the Legendre polynomials (eq. (2.14)) and after rearranging the terms, eq. (B.5) is rewritten to

$$r_E = \frac{\sum_n \frac{4n \hat{w}_{n-1} \hat{w}_n}{(2n-1)(2n+1)}}{\sum_n \frac{2}{(2n+1)} \hat{w}_n^2} = \frac{\hat{r}_E}{E}. \quad (\text{B.6})$$

This is akin to the 'max r_E ' optimization criterion presented in [DRP98] and [Dan00].

Maximization of r_E from eq. (B.6) is achieved by setting the derivatives with respect to \hat{w}_n to zero $\frac{\partial r_E}{\partial \hat{w}_n} \stackrel{!}{=} 0$, which leads to

$$\frac{n \hat{w}_{n-1}}{(2n-1)} + \frac{(n+1) \hat{w}_{n+1}}{(2n+3)} - r_E \hat{w}_n = 0.$$

And by substituting $\hat{w}_n = w_n(2n+1)$ the optimum weights are found from

$$n w_{n-1} + (n+1) w_{n+1} - r_E w_n (2n+1) = 0. \quad (\text{B.7})$$

Comparing eq. (B.7) with the recursion formula of the Legendre polynomials yields the ansatz $w_n = P_n(r_E)$. We may require the boundary value $w_{N+1} = P_{N+1}(r_E) = 0$ for the coefficients. As maximization is desired, only the highest zero of $P_{N+1}(r_E)$ meets the optimization criterion. Accordingly, the optimum order weights are defined as

$$w_n^{(opt)} = P_n(r_E) \text{ iff } r_E = \operatorname{argmax}_{r_E} \{P_{N+1}(r_E) = 0\}. \quad (\text{B.8})$$

The optimal order weights for the non-normalized Legendre polynomials are calculated as follows $\hat{w}_n^{(opt)} = w_n^{(opt)}(2n+1)$.

Bibliography

- [Abh99] P. Abhayapala, “Modal analysis and synthesis of broadband nearfield beamforming arrays,” Ph.D. dissertation, Citeseer, 1999.
- [Bau42] B. Bauer, “Super-cardioid directional microphone,” *Electronics*, *January*, 1942.
- [Bau11] R. Baumgartner, “Time domain fast-multipole translation for ambisonics,” M. Thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, 2011.
- [BCH08] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*. Springer Verlag, 2008, vol. 1.
- [BDM06] S. Bertet, J. Daniel, and S. Moreau, “3d sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone,” *120th Conv of the Audio Eng Soc*, vol. 5, 2006.
- [BPF11] R. Baumgartner, H. Pomberger, and M. Frank, “Practical implementation of radial filters for ambisonic recordings,” 2011, preprint (2011).
- [BR04] I. Balmages and B. Rafaely, “Room acoustics measurements by microphone arrays,” in *Electrical and Electronics Engineers in Israel, 2004. Proceedings. 2004 23rd IEEE Convention of*. IEEE, 2004, pp. 420–423.
- [BR07] —, “Open-sphere designs for spherical microphone arrays.” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 727–732, 2007. [Online]. Available: <http://dblp.uni-trier.de/db/journals/taslp/taslp15.html#BalmagesR07>

- [BSG⁺89] I. Bronstein, K. Semendjajew, G. Grosche, V. Ziegler, and D. Ziegler, *Taschenbuch der mathematik*. Deutsch, 1989.
- [BW01] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer Verlag, 2001.
- [CG77] P. Craven and M. Gerzon, “Coincident microphone simulation covering three dimensional space and yielding various directional outputs,” Aug. 16 1977, uS Patent 4,042,779.
- [CK07] E. Cheney and D. Kincaid, *Numerical mathematics and computing*. Brooks/Cole Pub Co, 2007.
- [Dan00] J. Daniel, “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia,” *Université Pierre et Marie Curie (Paris VI): Paris*, 2000.
- [DfR08] M. Dickreiter and S. für Rundfunktechnik, *Handbuch der Tonstudioteknik*. Saur, 2008, vol. 6.
- [DLZ⁺05] R. Duraiswami, Z. Li, D. Zotkin, E. Grassi, and N. Gumerov, “Plane-wave decomposition analysis for spherical microphone arrays,” in *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*. IEEE, 2005, pp. 150–153.
- [DRP98] J. Daniel, J. Rault, and J. Polack, “Ambisonics encoding of other audio formats for multiple listening conditions,” *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 1998.
- [EKM03] G. Elko, R. Kubli, and J. Meyer, “Audio system based on at least second-order eigenbeams,” Jul. 24 2003, wO Patent WO/2003/061,336.
- [EKM04] ———, “Audio system based on at least second; order eigenbeams,” Oct. 13 2004, eP Patent 1,466,498.
- [EKM⁺09] G. Elko, R. Kubli, J. Meyer *et al.*, “Audio system based on at least second-order eigenbeams,” Sep. 8 2009, uS Patent 7,587,054.
- [Far00] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” *Preprints-Audio Engineering Society*, 2000.

- [GD04] N. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier Science, 2004.
- [Gir96] F. Giron, “Investigations about the directivity of sound sources,” *Verlag Shaker, Aachen 1996*, pp. 103–113, 1996.
- [Hay85] S. Haykin, “Array signal processing,” *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 493 p. For individual items see A85-43961 to A85-43963.*, vol. 1, 1985.
- [HB04] Y. Huang and J. Benesty, *Audio signal processing for next-generation multimedia communication systems*. Springer, 2004.
- [HS96] R. Hardin and N. Sloane, “Mclaren’s improved snub cube and other new spherical designs in three dimensions,” *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, 1996.
- [JJ00] H. Jeffreys and B. Jeffreys, *Methods of mathematical physics*. Cambridge Univ Pr, 2000.
- [JW91] J. Jackson and J. Wiley, *A user’s guide to principal components*. Wiley Online Library, 1991.
- [KL80] V. Klema and A. Laub, “The singular value decomposition: Its computation and some applications,” *Automatic Control, IEEE Transactions on*, vol. 25, no. 2, pp. 164–176, 1980.
- [LR05] Z. Li and R. Duraiswami, “Hemispherical microphone arrays for sound capture and beamforming,” in *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*. IEEE, 2005, pp. 106–109.
- [ME02] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–1781.
- [ME08a] ———, “Spherical harmonic modal beamforming for an augmented circular microphone array,” in *Acoustics, Speech and Signal Processing, 2008*.

- ICASSP 2008. IEEE International Conference on.* IEEE, 2008, pp. 5280–5283.
- [ME08b] ———, “Augmented elliptical microphone array,” Jul. 9 2008, uS Patent App. 12/595,082.
- [MI86] P. Morse and K. Ingard, *Theoretical acoustics*. Princeton Univ Pr, 1986.
- [Mor06] S. Moreau, “Étude et réalisation d’outils avancés d’encodage spatial pour la technique de spatialisation sonore higher order ambisonics: microphone 3d et contrôle de distance,” Ph.D. dissertation, PhD thesis, Université du Maine, 2006.
- [Ple09] P. Plessas, “Rigid sphere microphone arrays for spatial recording and holography,” *BSc Thesis*, 2009.
- [PZ08] H. Pomberger and D. Zotter, “Angular and radial directivity control for spherical loudspeaker arrays,” *M. thesis, IEM, TU-Graz*, 2008.
- [Raf05] B. Rafaely, “Analysis and design of spherical microphone arrays,” *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 1, pp. 135–143, 2005.
- [Rei06] F. Reining, “Sound field microphone,” Dec. 27 2006, eP Patent 1,737,268.
- [Rei09] ———, “Microphone arrangement comprising pressure gradient transducers,” Feb. 23 2009, uS Patent App. 12/390,990.
- [RWB07] B. Rafaely, B. Weiss, and E. Bachmat, “Spatial aliasing in spherical microphone arrays,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 3, pp. 1003–1010, 2007.
- [SEA02] G. Stan, J. Embrechts, and D. Archambeau, “Comparison of different impulse response measurement techniques,” *Journal of the Audio Engineering Society*, vol. 50, no. 4, 2002.
- [SEH08] W. Song, W. Ellermeier, and J. Hald, “Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise,” *The Journal of the Acoustical Society of America*, vol. 123, p. 910, 2008.

- [SW88] M. Sessler and J. West, “Unidirectional second order gradient microphone,” May 3 1988, uS Patent 4,742,548.
- [Teu06] H. Teutsch, “Wavefield decomposition using microphone arrays and its application to acoustic scene analysis,” Ph.D. dissertation, Universitätsbibliothek, 2006.
- [TF01] A. Torger and A. Farina, “Real-time partitioned convolution for ambiphonics surround sound,” in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 195–198.
- [VT02] H. Van Trees, *Optimum array processing*. Wiley Online Library, 2002.
- [VVB88] B. Van Veen and K. Buckley, “Beamforming: A versatile approach to spatial filtering,” *ASSP Magazine, IEEE*, vol. 5, no. 2, pp. 4–24, 1988.
- [WA01] D. Ward and T. Abhayapala, “Reproduction of a plane-wave sound field using an array of loudspeakers,” *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 6, pp. 697–707, 2001.
- [Wil99] E. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. Academic Pr, 1999.
- [YSS⁺11] S. Yan, H. Sun, U. Svensson, X. Ma, and J. Hovem, “Optimal modal beamforming for spherical microphone arrays,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 361–371, 2011.
- [ZN07] F. Zotter and M. Noisternig, “Near-and farfield beamforming using spherical loudspeaker arrays,” in *Proceedings of the 3rd Congress of the Alps Adria Acoustics Association, Graz, Austria*, vol. 2007, 2007.
- [Zot08] F. Zotter, “Sampling strategies for acoustic holography/holophony on the sphere,” *Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria., Tech. Rep*, pp. 2–4, 2008.
- [Zot09] —, “Analysis and synthesis of sound-radiation with spherical arrays,” Ph.D. dissertation, IEM-Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria, 2009.

- [ZPF09] F. Zotter, H. Pomberger, and M. Frank, “An alternative ambisonics formulation: Modal source strength matching and the effect of spatial aliasing,” *Proc. of the 126th AES Conv., Munich*, 2009.
- [ZRSH03] J. Zmólnig, W. Ritsch, A. Sontacchi, and R. Höldrich, “The iem cube,” in *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, 2003, pp. 127–130.