Diploma Thesis

Phase Unwrapping for Spherical Interpolation of Head-Related Transfer Functions

Revised Version

Johannes Zaar

Assessor: Robert Höldrich

Supervisors: Franz Zotter

Markus Noisternig

Institute of Electronic Music and Acoustics
University of Music and Performing Arts, Graz
Institut de Recherche et Coordination Acoustique/Musique, Paris

Graz, December 15, 2011

Acknowledgements

Many people have been helpfully involved in my work on this thesis.

First of all, I want to thank the supervisors of this thesis:
Franz Zotter for all the time, enthusiasm, and his never ending stream of ideas he contributed whenever he could. These ideas and the resulting discussions always led to some kind of progress or deeper understanding.
Markus Noisternig for his systematic thinking and structuring and the ability and willingness to come up with important questions. Further, for his general helpfulness concerning not only the scientific, but also the logistic problems I had to solve during my stay in Paris.

I also like to thank Piotr Majdak for his advice on literature and for the interesting discussions, Thibaut Carpentier for his support concerning technical issues and measurement data, and Brian Katz for the opportunity to simulate HRTFs using the BEM-model at LIMSI/CNRS. Further, I want to thank Hannes Pomberger for his code for the visualization of spherical harmonics and Matthias Frank for his advice on the psychoacoustic evaluation of the results.

I am very grateful for the opportunity to carry out part of this work at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris. In this context I want to thank the IRCAM, the Institute of Electronic Music and Acoustics, the Department of International Relations of the University of Music and Performing Arts, Graz, and the Österreichischer Austauschdienst (OEAD).

A lot of people helped me to deal with the unfamiliar situation in Paris on a personal level. To mention just a few, thank you Julien, José, Sarah, Kevin, Léa, Rémi.

Thank you Klotzi, Joschi and Sebastian for the music and everything else, it's really way too much to write it down here.

Thank you Miranda.

Thanks to my family for always having my back.

Abstract

Individual head-related transfer functions (HRTFs) allow for an accurate representation of three-dimensional audio scenes via headphones. Given the high spatial resolution necessary for good localization results, a set of HRTFs usually comprises a large number of impulse responses (FIR filters), one filter pair for every measured source direction. This leads to a very long measurement duration, because every acoustic path has to be measured separately. Furthermore, the resulting large sets of filters require a high memory capacity. A more compact representation using high-quality interpolation of HRTFs is hence desirable. Interpolation of HRTFs is typically performed using a decomposition into onset delays and minimum phase components prior to the interpolation itself. In this thesis, an interpolation of unmodified HRTFs is investigated, using an adapted discrete spherical harmonic transform. In doing so, suitable HRTF features that are to be interpolated, e.g. time-delay and attenuation, deserve some notice as to avoid undesired artifacts. Especially the unwrapped phase, which can be considered as a frequency-dependent temporal delay, is intensely investigated. As a result of this investigation, a spherical phase unwrapping algorithm based on the concept of neighbouring points is presented. The results are verified via psychoacoustical error measures using measured and simulated HRTFs.

Zusammenfassung

Individuelle kopfbezogene Übertragungsfunktionen (HRTFs) ermöglichen eine sehr genaue Wiedergabe von dreidimensionalen Schallfeldern über Kopfhörer. Aufgrund der hohen räumlichen Auflösung, die für gute Lokalisationsergebnisse nötig ist, umfasst ein Satz von HRTFs üblicherweise eine hohe Anzahl an Impulsantworten (FIR-Filter), ein Paar für jede gemessene Richtung. Dies führt zu einer sehr langen Messdauer, da jeder akustische Übertragungspfad separat gemessen werden muss. Zudem benötigen die resultierenden Filtersätze viel Speicherplatz. Eine kompaktere Darstellung mittels hochqualitativer Interpolation von HRTFs ist daher wünschenswert. HRTFs werden üblicherweise interpoliert, indem vor der Interpolation selbst eine Zerlegung in Laufzeiten und minimalphasige Komponenten durchgeführt wird. In dieser Arbeit wird die Interpolation von unveränderten HRTFs mittels einer angepassten diskreten Spherical Harmonic Transformation untersucht. Dabei bedürfen zu interpolierende HRTF-Eigenschaften, z.B. zeitliche Verzögerung und Abdämpfung, besonderer Beachtung um unerwünschte Artefakte zu vermeiden. Insbesondere die abgerollte '*unwrapped*' Phase, welche als frequenzabhängige zeitliche Verzögerung betrachtet werden kann, wird intensiv untersucht. Als Resultat dieser Untersuchung wird ein Algorithmus für Phase Unwrapping über der Kugel präsentiert, der auf dem Konzept der geometrischen Nachbarschaft basiert. Die Ergebnisse werden mittels psychoakustischer Fehlermaße für simulierte und gemessene HRTFs überprüft.

Contents

1 Introduction **6**

2 Head-Related Transfer Functions **8**

2.1 Basics 8

2.2 Coordinate Systems 9

2.3 Interpolation Approaches 10

2.3.1 Common Approach Neglecting All-Pass Components 11

2.3.2 Approach Including all Components 13

2.4 Evans' Results 16

3 Discrete Spherical Harmonic Transform and Interpolation **19**

3.1 Basic Definition 19

3.2 Symmetry Characteristics 20

3.3 Discrete Spherical Harmonic Transform 22

3.4 Reconstruction and Interpolation 23

3.5 Invertibility and Regularization 23

3.6 Definition of Maximal Order for Spatial Sampling Grids 24

3.7 Aliasing 26

3.8 Reconstruction and Interpolation of Simulated HRTFs 27

3.8.1 Reconstruction and Interpolation Quality 29

3.8.2 Most Important Basis Functions 31

4 Phase Unwrapping **33**

4.1 Spatial Phase Distribution 33

4.2 The Concept of Phase Unwrapping 36
4.3 Spherical Phase Unwrapping 38
4.3.1 Concept of Neighboring Points 38
4.3.2 Algorithm 40
4.3.3 Boundaries 43

5 Spectral vs. Spherical Phase Unwrapping 46
5.1 Spatial Continuity 46
5.2 Spherical Harmonic Transform Performance 51

6 Results 54
6.1 ITD Error 54
6.2 ILD Error 57
6.3 Magnitude Error 59
6.4 Most Important Basis Functions 61

7 Conclusion 65
7.1 Summary of the results 65
7.2 Interpretation and Outlook 66

8 Annex 73
8.1 Minimum Phase Calculation 73
8.2 Onset Delay Extraction 74
8.3 Onset Delay Simulation 74

Chapter 1

Introduction

The initial motivation for the work documented in this thesis was the idea of finding an ideal representation of the head-related transfer functions (HRTFs) in space. Since HRTFs are very complex spatial filters, they have to be measured with a high spatial resolution, i.e. for many spherically distributed sampling positions. *Ideal* in this context hence refers to a compact and spatially continuous representation that facilitates accurate interpolation using as few spatial sampling positions as possible. This can be useful concerning a possible reduction of the necessary measurement sampling grid, thus reducing the measurement duration, which is typically between 30 and 60 minutes. Moreover, a compact representation of HRTFs is beneficial in terms of computing power, particularly when the filtering is to be applied in real-time.
Partly based on the work of Evans et al. [EAT98], who investigated the performance of the discrete spherical harmonic transform (DSHT) concerning the interpolation of different HRTF features, a basic concept was established. The human auditory system localizes sound sources using the spectral shaping caused by the head and body and the frequency-dependent propagation delay. An interpolation of the HRTFs' magnitude responses and group delays should therefore be well-adapted to the human perception. A compact and spatially continuous representation of the transfer functions measured at discrete locations can be realized using DSHT. Slight symmetry properties of the spatial group delay distribution can be enforced utilizing symmetries of the spherical harmonic functions. The alignment of the respective symmetry properties is achieved by rotating the basis functions, which is meant to support a compact representation or manipulation referring to symmetry.
The group delay of the HRTFs is not directly accessible and its calculation is very prone to errors. The frequency-dependent propagation delay can also be represented by the absolute phase. As only the wrapped phase is accessible, phase unwrapping has to be applied to calculate the absolute phase. Phase unwrapping is a very delicate procedure and will hence be intensely investigated. Phase unwrapping over the sphere will be presented and compared to the typically employed spectral phase unwrapping.

The structure of the thesis is summarized as follows.
chapter 2 introduces the head-related transfer functions, discusses their features concerning their significance for human sound source localization and defines the employed coordinate system. Further, the commonly used minimum phase and delay approximation of HRTFs is discussed and the interpolation approach using unmodified HRTFs is investigated. Additionally, a review of the related work of Evans et al. is provided.
Chapter 3 presents the adapted discrete spherical harmonic transform. This requires an investigation of the ideal band limitation and a verification of the assumptions about the suitability of different HRTF representations for the DSHT using simple simulated filters.
Chapter 4 contains an investigation of spectral phase unwrapping with regard to its impacts on the spatial phase continuity. The theory of phase unwrapping is explained in general and a spherical approach to phase unwrapping is derived and briefly discussed.
Chapter 5 compares spectral and spherical phase unwrapping with regard to spatial phase continuity. Furthermore, the suitability of the resulting unwrapped phase responses for the DSHT is discussed.
In chapter 6, the suitability of different HRTF representations for the DSHT is evaluated by means of psychoacoustically motivated error measures. Further, the spherical harmonics required for the investigated HRTF representations are compared.
Chapter 7 concludes the thesis by summarizing and interpreting the results and provides an outlook to further work on the issues discussed in this thesis.

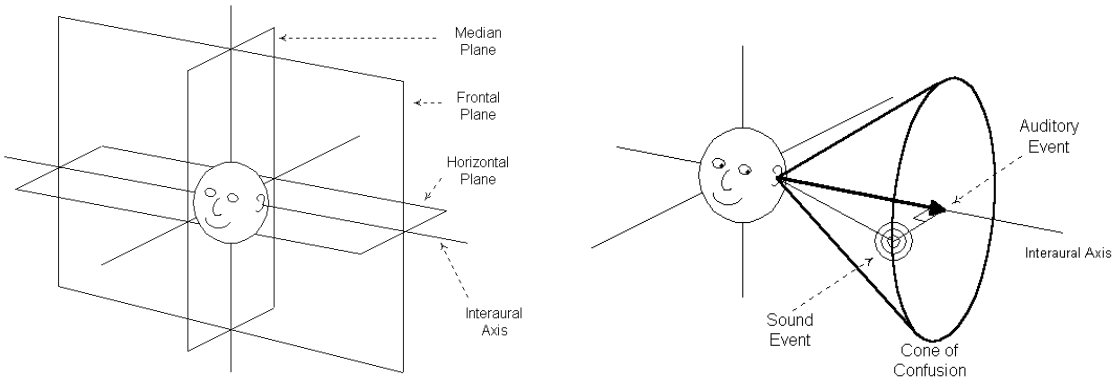
Chapter 2

Head-Related Transfer Functions

2.1 Basics

Localization cues. Research on human sound source localization suggests that the human auditory system uses different types of primary localization cues. According to the classic 'duplex theory' [MM02], the lateral position (see *figure 2.2.1*) is determined using the interaural time difference (ITD) for frequencies below $1.3\mathrm{kHz}$ and using the interaural level difference (ILD) for frequencies above $1.3\mathrm{kHz}$. The ITD represents the propagation delay between the ears and can only be detected without ambiguity if the wavelength λ is smaller than the length of the acoustic path between the ears. Assuming a head diameter of $17\mathrm{cm}$, this results in a maximal frequency of $\sim 1.3\mathrm{kHz}$. The ILD represents frequency-dependent level differences between the ears, which result from shadowing effects for smaller wavelengths caused by the head. The binaural ITD and ILD cues do not provide information concerning the elevation of a sound source since only the lateral position is detected. Consequently, the detected position could lie anywhere in the respective sagittal plane, a plane parallel to the median plane (see *figure 2.1.1(a)*). This ambiguity is known as the 'cone of confusion', depicted in *figure 2.1.1(b)*. The filtering effect of the pinna is a disambiguating localization cue, mostly in the frequency-region between $4\mathrm{kHz}$ and $8\mathrm{kHz}$, which provides information about the position of a sound source on the cone of confusion and is particularly important in the median plane, where ITD and ILD are zero.

A problem that remains is the front-back-ambiguity, which has been reported in psychoacoustic studies using static filtering. Wightman and Kistler investigated this phenomenon in [FLW99]. Results suggest that humans resolve these ambiguities with small head movements in reality. This can only be reproduced by using interaction of head movements with suitable time-variant filtering.



(a) Planes and axes (b) Cone of confusion

Figure 2.1.1: Human sound source localization: Spatial definitions [Wes]

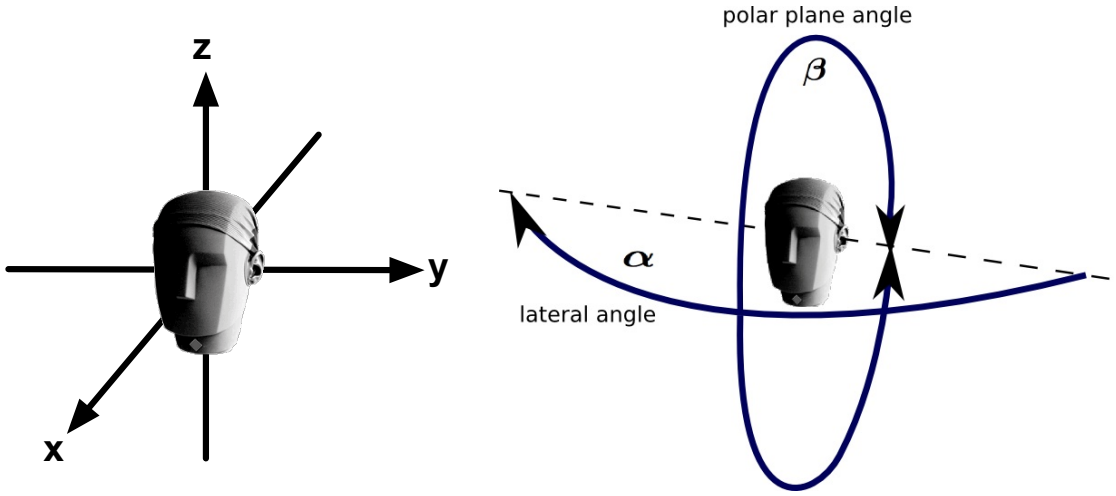
HRTFs and HRIRs. The filtering characteristics of the human head and body can be simulated by a set of filters measured on a dense sampling grid on a surrounding sphere. These filters are called head-related transfer functions (HRTFs) or head-related impulse responses (HRIRs). HRTFs are functions of the direction vector $\boldsymbol{\theta}$ and the frequency, denoted as $H(\boldsymbol{\theta}, f)$, while HRIRs depend on position and time, denoted as $h(\boldsymbol{\theta}, t)$. The distance of sound sources is usually neglected for the commonly used far-field HRTFs, which have to be measured at a minimum distance of $1\mathrm{m}$ to the center of the head. Independent of frequency or time domain representation, the head-related filters are called HRTFs in the following. When both terms are used, HRTFs refer to the frequency domain representation and HRIRs to the time domain representation.

2.2 Coordinate Systems

Since the spatial representation, properties and manipulation of HRTFs is the key interest of this thesis, the three-dimensional spherical coordinate system that will be used throughout the thesis is introduced in this section.

Cartesian coordinate system. A typical definition of a subjective Cartesian coordinate system is right handed, with the x-axis pointing in the viewing direction and the y-axis corresponding to the interaural axis, see *figure 2.2.1(a)*.

Interaural polar coordinate system. The interaural polar coordinate system fits well with the human perception of sound sources in space, and is thus often used in psychoacoustic considerations. Every angular position is expressed via two angles. The angles reflect the spatial descent of the different localization cues: The ITD and ILD depend on the lateral



(a) Cartesian (b) Interaural polar

Figure 2.2.1: Coordinate systems

angle α, while the spectral cues depend mainly on the polar plane angle β, see *figure 2.2.1(b)*. The conversion from Cartesian to interaural polar coordinates is perfomed as follows

$$\begin{aligned} \alpha &= \arctan\left(\frac{\sqrt{x^2+z^2}}{y}\right), \\ \beta &= \arctan\left(\frac{z}{x}\right), \\ r &= \sqrt{x^2+y^2+z^2}. \end{aligned} \tag{2.2.1}$$

Inversely, the conversion from interaural polar to Cartesian coordinates can be written as

$$\begin{aligned} x &= r\ \sin\alpha\ \cos\beta, \\ y &= r\ \cos\alpha, \\ z &= r\ \sin\alpha\ \sin\beta. \end{aligned} \tag{2.2.2}$$

The distance variable r will be neglected in the following.

2.3 Interpolation Approaches

The interpolation of HRTFs has been widely discussed in literature in terms of its psychoacoustical limits, which naturally have physical reasons since the human auditory system is restricted to the available acoustic information. This raises the question, how many measurement points are necessary to generate the impression of an acoustic real-world situation, which has been

investigated by Minnaar et al. [MPC05]. The accuracy of the interpolation strongly depends on the distribution of the sampling positions and on the quality of the interpolation procedure. There are two main approaches to the spatial interpolation of HRTFs. The *local approaches* generate an interpolation between the adjacent filters of the desired position, while the *global approaches*, on the other hand, take all filters into account.

The following paragraphs focus on the question, which representation of HRTFs is most suitable for interpolation in general. A brief example is demonstrated dealing with a local interpolation between two filters. A similar behaviour can be assumed for the global interpolation via discrete spherical harmonic transform, the approach that is the matter of the investigations.

2.3.1 Common Approach Neglecting All-Pass Components

A very common method for the interpolation and real-time processing of HRTF filters is based on the assumption that they can be sufficiently approximated by their minimum phase counterpart and an onset delay, which is often referred to as *excess phase* in literature. This assumption has largely been verified in psychoacoustic experiments [KIC99, TM08] and via simulation [NKA08]. Prior to the interpolation itself, a decomposition of the HRTFs is performed, which manipulates the phase response of the filters and thus simplifies their structure.

HRTF decomposition. A discrete complex HRTF spectrum $H[\omega_l]$ can be split into magnitude $|H[\omega_l]|$ and phase $\phi[\omega_l]$

$$H[\omega_l] = |H[\omega_l]|\ \mathrm{e}^{\mathrm{j}\phi[\omega_l]}, \qquad \phi[\omega_l] = \arctan\left(\frac{\Im\{H[\omega_l]\}}{\Re\{H[\omega_l]\}}\right),$$

where $\omega_l = 2\pi\frac{l}{\mathrm{L}}\mathrm{f_s}$ (with sampling rate $\mathrm{f_s}$, the length of the discrete Fourier transform L and the discrete frequency bins $l = 0,\ 1,\ ...,\ \mathrm{L}-1$) and $\Im$ and $\Re$ denote the imaginary and real part, respectively. The phase can be further decomposed into minimum phase and excess phase (see annex, chapter 8.1, for the calculation of the minimum phase)

$$H[\omega_l] = |H[\omega_l]|\ \mathrm{e}^{\mathrm{j}\phi_{\mathrm{min}}[\omega_l]}\ \mathrm{e}^{\mathrm{j}\phi_{\mathrm{excess}}[\omega_l]},$$

where $\phi_{\mathrm{excess}}[\omega_l] = \phi[\omega_l] - \phi_{\mathrm{min}}[\omega_l]$. The assumption for the simplification of the filters is

$$\phi_{\mathrm{excess}}[\omega_l] = \phi_{\mathrm{lin}}[\omega_l], \tag{2.3.1}$$

where the linear phase

$$\phi_{\mathrm{lin}}[\omega_l] = -\omega_l\ \tau$$

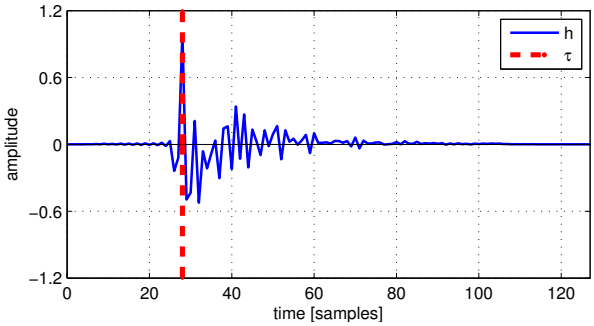
represents the temporal onset delay τ of the corresponding impulse response $h[k]$ (see annex, chapter 8.2, for the calculation of τ). The modified HRIR $h_{\mathrm{mod}}[k]$ can now be represented by its minimum phase version

$$h_{\mathrm{min}}[k] = IDFT\{|H[\omega_l]|\ \mathrm{e}^{\mathrm{j}\phi_{\mathrm{min}}[\omega_l]}\}$$

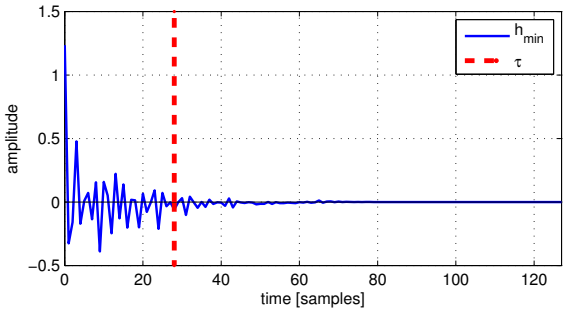
and the onset delay τ, in mathematical terms a convolution of $h_{\mathrm{min}}[k]$ with a unit impulse delayed by τ in samples

$$h_{\mathrm{mod}}[k] = \delta[k-\tau] * h_{\mathrm{min}}[k],$$

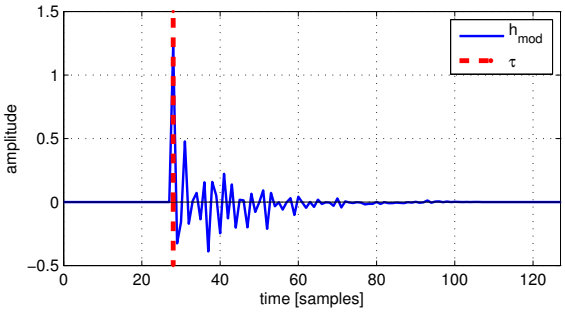
where k denotes the discrete temporal samples.



(a) Original impulse response h and onset delay τ



(b) Minimum phase impulse response h_{min} and onset delay τ



(c) Resulting modified impulse response h_{mod} and onset delay τ

Figure 2.3.1: Minimum phase and delay decomposition of an HRIR

The spatial interpolation is applied to the delay τ and the minimum phase HRIRs separately. The interpolation result is then obtained similarly by convolving the interpolated time-shift with the interpolated minimum phase HRIR.

Advantages. This decomposition of the HRIRs leads to advantages concerning the interpolation and also the real-time rendering of HRIRs, because the minimum phase versions provide *a) perfect temporal alignment*, which is crucial for accurate time domain interpolation, and *b) the least possible amount of important filter samples*, which facilitates a reduction of the

filter length and thus reduces the computational complexity.

Drawbacks. Since HRTFs are in fact mixed-phase systems, i.e. *non-minimum phase in nature* [Bus06, NKA08], the basic assumption from *equation 2.3.1* concerning the excess phase cannot be verified [NKA08]. The original phase typically splits into three parts

$$\phi[\omega_l] = \phi_{\mathrm{min}}[\omega_l] + \phi_{\mathrm{lin}}[\omega_l] + \phi_{\mathrm{ap}}[\omega_l],$$

where $\phi_{\mathrm{ap}}[\omega_l]$ is the all-pass phase that has been neglected above. Provided the extracted onset delay matches the low-frequency group delay, the modification of the phase response does not have significant perceptual consequences [KIC99, TM08]. Nevertheless, the *detection or simulation of the onset delay* [BNK05] is a very delicate task, which requires maximal accuracy since errors directly influence the $ITD = \tau_{\mathrm{right}} - \tau_{\mathrm{left}}$, leading to lateral changes in the perceived direction. Moreover, the temporal resolution depends on the sampling frequency in time-discrete signal processing, e.g. $23\mu\mathrm{s}$ at $\mathrm{f_s} = 44.1\mathrm{kHz}$, which results in small ITD errors due to rounding.

2.3.2 Approach Including all Components

In this thesis, an alternative method for the interpolation of HRIRs is investigated, using the unmodified, original impulse responses. This section compares the interpolation accuracy resulting from the interpolation of different filter representations.

Depending on the distance between the source position and the ear, HRTF-filters have different onset delays and peak levels. In order to clarify the basic interpolation characteristics, two simple impulse responses $h_1[k]$ and $h_2[k]$ are considered (*figure 2.3.2*). $h_2[k] = 0.6\ \delta(k-4) * h_1[k]$ is an attenuated version of $h_1[k]$, temporally shifted by 4 samples.

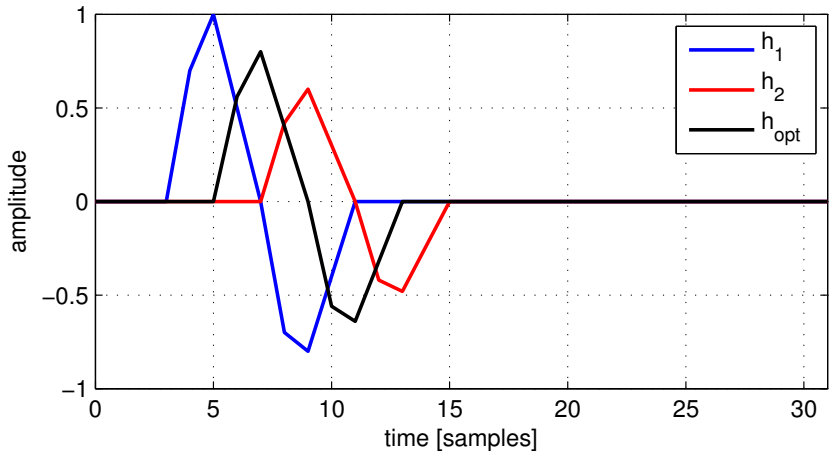


Figure 2.3.2: Exemplary impulse responses

The optimal interpolation result is defined as the mean of amplitude and temporal displacement, $h_{\mathrm{opt}}[k] = 0.8\ \delta[k-2] * h_1[k]$.

Direct interpolation of HRIRs. A very simple approach is to directly interpolate the impulse responses. In mathematical terms, the interpolated impulse response $h_{\mathrm{int}}[k]$ is calculated as the mean of $h_1[k]$ and $h_2[k]$

$$h_{\mathrm{int}}[k] = \frac{h_1[k] + h_2[k]}{2}.$$

The problem of this interpolation technique becomes quite obvious in *figure 2.3.3*. Unless the impulse responses are temporally aligned, which cannot be assumed for HRIRs, the interpolation crossfades the adjacent peaks instead of shifting them in time and adjusting their levels, leading to comb filters. Although $h_1[k]$ and $h_2[k]$ are quite unrealistic examples in terms of the temporal displacement, this example shows that the direct interpolation method is not adequate for HRIRs.

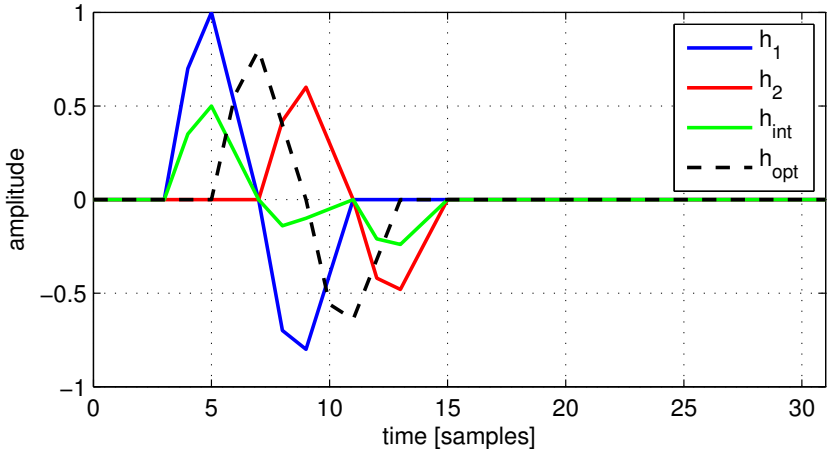


Figure 2.3.3: Result of direct HRIR interpolation

Interpolation of unwrapped phase and magnitude. A very different approach is an interpolation of the HRTFs using a spectral representation. More precisely, the spectrum is decomposed into magnitude $|H[\omega_l]|$ and unwrapped phase $\Phi[\omega_l]$

$$H[\omega_l] = |H[\omega_l]|\ \mathrm{e}^{\mathrm{j}\,\Phi[\omega_l]},$$

where Φ denotes the unwrapped version of the phase angle ϕ,

$$\Phi[\omega_l] = 2\pi\eta + \phi[\omega_l].$$

$\eta \in \mathbb{Z}$ is an integer variable defined according to the constraint that the absolute difference between two adjacent frequency bins must not be greater than π

$$|\Phi[\omega_l] - \Phi[\omega_{l-1}]| \leq \pi.$$

The phase angle $\phi[\omega_l] = \arctan\left(\frac{\Im\{H[\omega_l]\}}{\Re\{H[\omega_l]\}}\right)$, which is also referred to as wrapped phase, can be derived from the spectrum. It lies in the interval $]-\pi, \pi]$ and typically represents a discontinuous function of frequency. By contrast, the unwrapped phase is continuously distributed over frequency. The wrapped phase can be calculated from the unwrapped phase via the phase-wrapping operator $\mathcal{W}$, also known as the principle phase argument function:

$$\begin{aligned} \phi &= \mathcal{W}(\Phi) \\ &= \mathrm{mod}_{2\pi}(\Phi + \pi) - \pi. \end{aligned} \tag{2.3.2}$$

The unwrapped phase is related to the group delay τ_{G} via integration

$$\Phi(\omega) = -\int \tau_{\mathrm{G}}(\omega)\mathrm{d}\omega \quad \Longleftrightarrow \quad \Phi[\omega_l] = -\frac{2\pi}{\mathrm{L}} \sum_{i=0}^{l} \tau_{\mathrm{G}}[\omega_i] \ ,$$

where $\omega = 2\pi f$. The inverse transformation writes as

$$\tau_{\mathrm{G}}(\omega) = -\frac{\partial \Phi(\omega)}{\partial \omega} \quad \Longleftrightarrow \quad \tau_{\mathrm{G}}[\omega_l] = -\left[\Phi[\omega_l] - \Phi[\omega_{l-1}]\right] \frac{\mathrm{L}}{2\pi}.$$

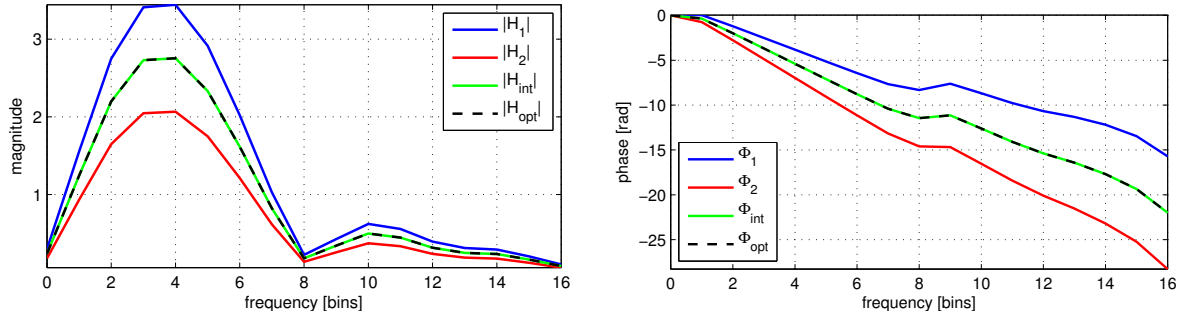
The unwrapped phase can be considered as a frequency dependent temporal measure similar to the group delay. That leaves us with a *representation of the HRTFs as magnitude* $|H[\omega_l]|$ *and 'temporal delay'* $\Phi[\omega_l]$ *for every discrete frequency* ω_l *of the spectrum*. The interpolation can now be applied using the discrete spectra of the impulse responses, $H_1[\omega_l]$ and $H_2[\omega_l]$, and separately interpolating magnitude and unwrapped phase values

$$\begin{aligned} |H_{\mathrm{int}}[\omega_l]| &= \frac{|H_1[\omega_l]| + |H_2[\omega_l]|}{2}, \\ \Phi_{\mathrm{int}}[\omega_l] &= \frac{\Phi_1[\omega_l] + \Phi_2[\omega_l]}{2}. \end{aligned}$$

The original and interpolated magnitude and phase responses are shown in *figure 2.3.4*; obviously, the interpolated versions $|H_{\mathrm{int}}[\omega_l]|$ and $\Phi_{\mathrm{int}}[\omega_l]$ match the optimal solutions $|H_{\mathrm{opt}}[\omega_l]|$ and $\Phi_{\mathrm{opt}}[\omega_l]$.

Using interpolated magnitude $|H_{\mathrm{int}}|$ and interpolated unwrapped phase Φ_{int}, the resulting filter impulse response can now be calculated by the inverse discrete fourier transform $(IDFT)$

$$h_{\mathrm{int}}[k] = IDFT\{|H_{\mathrm{int}}[\omega_l]| \ \mathrm{e}^{\mathrm{j}\,\Phi_{\mathrm{int}}[\omega_l]}\}. \tag{2.3.3}$$



(a) Magnitude $|H[\omega_l]|$ (b) Unwrapped phase $\Phi[\omega_l]$

Figure 2.3.4: Interpolation of magnitude and unwrapped phase

As can be observed in *figure 2.3.5*, the interpolation of the magnitude and unwrapped phase representation seems to achieve a very accurate result. The onset of the resulting impulse response $h_{\text{int}}[k]$ lies exactly between those of $h_1[k]$ and $h_2[k]$, while the shape is the same that an interpolation of temporally aligned filters would provide. It therefore exactly matches the previously defined optimal interpolation solution $h_{\text{opt}}[k]$. Hence, the magnitude and unwrapped phase approach clearly outperforms the direct interpolation of HRIRs without needing any kind of prior modification of the filters, as opposed to the minimum phase and delay approach. Note that this interpolation assumes that time shifting and magnitude interpolation is the most probable that will happen. It will not cover the case when spectral valleys or peaks are moving with respect frequency.

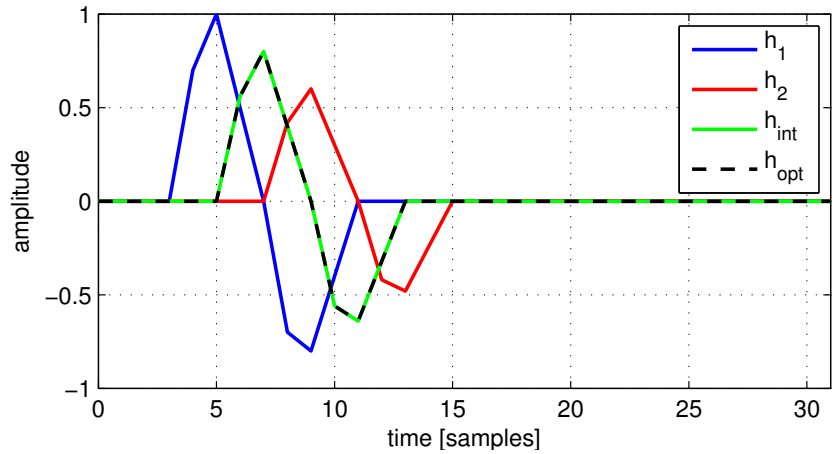


Figure 2.3.5: Resulting impulse response of magnitude and unwrapped phase interpolation

2.4 Evans' Results

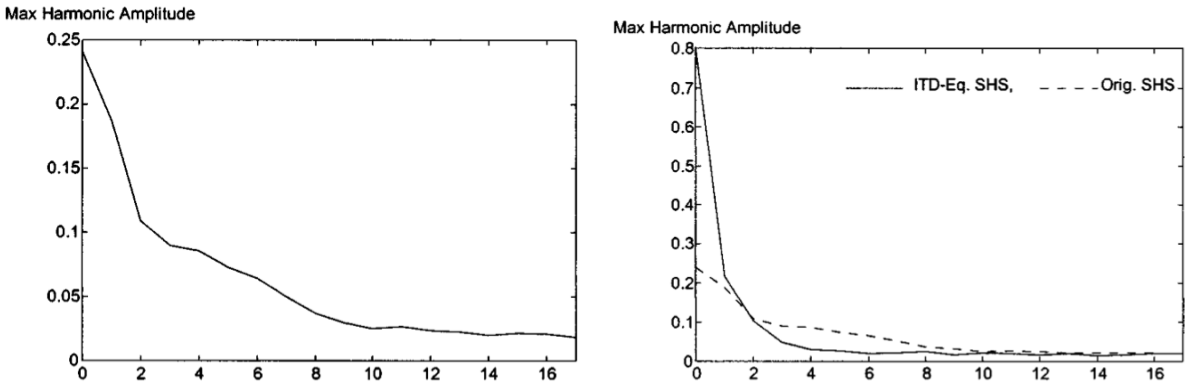
Global interpolation of HRTFs can be performed using the discrete spherical harmonic transform. Further information on the discrete spherical harmonic transform and the corresponding

nomenclature is given in chapter 3.

Evans et al. analyzed the reconstruction and interpolation performance of the discrete spherical harmonic transform of HRTFs [EAT98]. They used dummy head measurements on a spherical sampling grid with $10°$-spacing in azimuth and the corresponding elevation angles resulting from Gaussian quadrature. The HRTFs were sampled with $\mathrm{f_s} = 20.2\mathrm{kHz}$, leading to a maximal representable frequency of $\mathrm{f_s}/2 = 10.1\mathrm{kHz}$. Three different types of HRTF representations were used, 1) HRIRs, 2) ITD-equalized HRIRs and 3) HRTF magnitude and unwrapped phase. The ITD-equalized HRIRs are the original HRIRs without the onset delay. The maximal spherical harmonic order was set to $\mathrm{N} = 17$.

The results suggest that magnitude and unwrapped phase are best suited for an accurate representation of HRTFs using the discrete spherical harmonic transform. The article also contains some interesting results concerning the most important spherical harmonic orders. They evaluated this by defining a spherical harmonic spectrum (SHS) as the mean energy contained within the components of each order. The maximum values over time were used for the SHS of HRIRs and ITD-equalized HRIRs and the maximum values over frequency for the SHSs of magnitude and unwrapped phase.

While the SHS of the HRIRs in *figure 2.4.1*(a) reflects a rather large importance of the high-order components, the SHS of the ITD-equalized HRIRs in *figure 2.4.1*(b) decreases very steeply. Since the significance of high-order components is proportional to the complexity of the spherical pattern, this seems logical considering the temporal alignment of the ITD-equalized HRIRs.



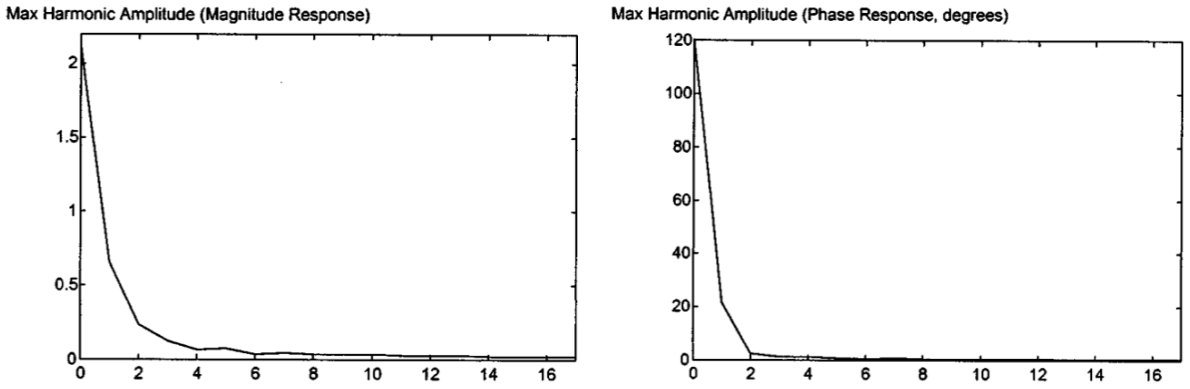
(a) Maxima of impulse responses over time

(b) Maxima of ITD-equalized impulse responses over time

Figure 2.4.1: Spherical harmonic spectra from Evans et al. [EAT98]

The SHS of the HRTF magnitude in *figure 2.4.2*(a) also seems surprisingly steep. The crucial information seems to be contained in the basis functions up to order 5; considering the importance of spectral cues for the fine localization on the cone of confusion, one may expect a more complex spherical pattern resulting in higher values of high-order components. The SHS

of the unwrapped phase in *figure 2.4.2*(b) even reflects a very low significance of components with a higher order than 1.



(a) Maxima of magnitude over frequency (b) Maxima of unwrapped phase over frequency

Figure 2.4.2: Spherical harmonic spectra from Evans et al. [EAT98]

These results lead to the question up to which order the spherical harmonics significantly contribute to an accurate representation of HRTFs. The maximal spherical harmonic order might be reducible to $\mathrm{N} = 5$ in this case without a significant loss of information, given an appropriate choice of the HRTF representation.

Chapter 3

Discrete Spherical Harmonic Transform and Interpolation

Spherical harmonic basis functions are well-defined orthonormal continuous functions in space which provide good interpolation characteristics. Every square integrable function defined over the spherical angles can be decomposed into a series of spherical harmonic coefficients. Most physical patterns known at many discrete positions fulfill this condition. Such a pattern can be decomposed into a weighted sum of spherical harmonics and thus be converted into a spatially continuous form. The discrete spherical harmonic transform (DSHT) has been widely used for the representation and simulation of HRTFs, e.g. in [EAT98, GODZ10, ZDG09, NSMH03, DZG04, NCNW10].

3.1 Basic Definition

The real-valued spherical harmonics can be defined as (see [Pom08])

$$Y_n^m(\beta,\alpha) = \begin{cases} \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\alpha)\sqrt{2-\delta_{m0}}\cos(m\beta), & \text{if } m \geq 0 \\ \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\alpha)\sqrt{2}\sin(m\beta), & \text{if } m < 0 \end{cases} \tag{3.1.1}$$

where the *spherical harmonic order* $n = 0, 1, 2, ..., \mathrm{N}-1, \mathrm{N}$ (N being the maximal order), the *spherical harmonic degree* $m = -n, -n+1, ..., n-1, n$ and P_n^m denotes the *associated Legendre functions* consistent with those in [Pom08].

The definition of the angles β and α (see *equation 2.2.1*) is different from the commonly used azimuth and elevation angles, e.g. in [Wil99]. By using the polar plane angle β and the lateral angle α, the spherical harmonics are not defined as usual but rotated by $90°$. This re-definition

is not degrading the accuracy of the transform since representing continuous functions in terms of order limited spherical harmonics is rotation invariant. As will be shown in chapter 3.2, this definition corresponds to an alignment of the symmetry characteristics of HRTFs and the spherical harmonics.

Figure 3.1.1 depicts the first 16 real-valued spherical harmonics. The order n is plotted vertically, increasing from top to bottom, while the degree m is plotted horizontally, from $-n$ on the left to n on the right. For every order n, $2n+1$ spherical harmonics exist, adding up to $(\mathrm{N}+1)^2$.
Two observations can be made here concerning the shape, 1) the spatial complexity is directly proportional to the order n and 2) all spherical harmonics of degree $m = 0$ are rotationally symmetric to the y-axis.

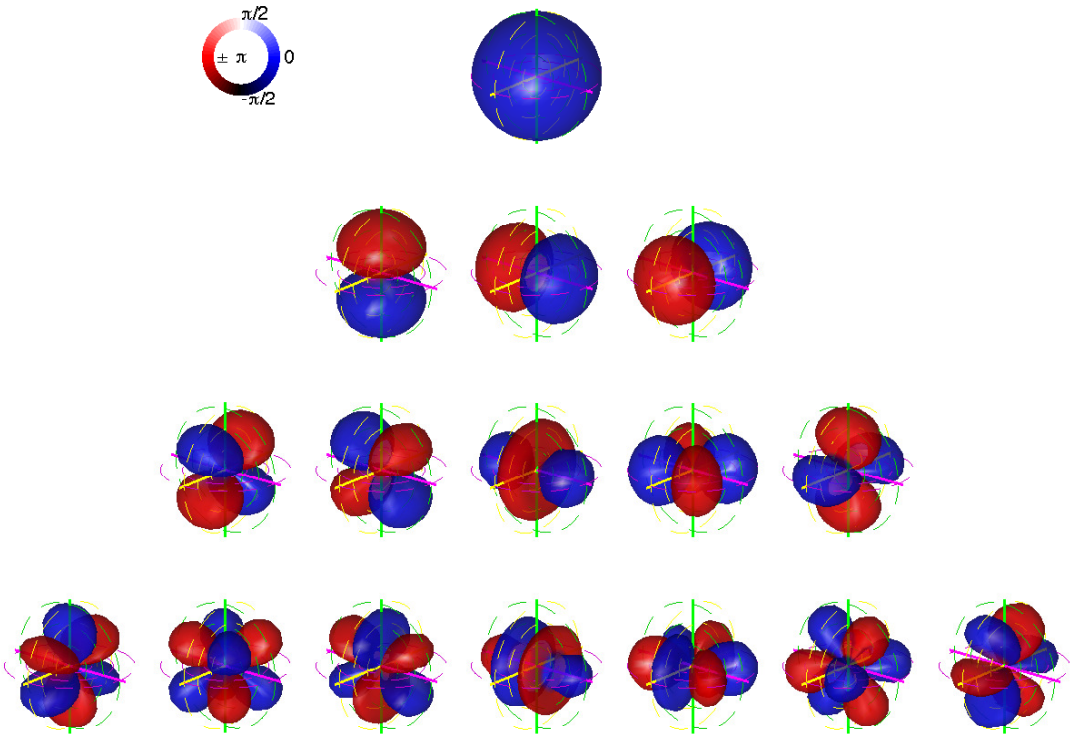


Figure 3.1.1: Spherical harmonic basis functions up to order 3, rotated by $90°$. x-axis: yellow, y-axis: pink, z-axis: green; plot functions by courtesy of Hannes Pomberger [Pom08]

3.2 Symmetry Characteristics

Considering the unwrapped phase as a temporal measure and the human head as approximately spherical, the phase distribution of HRTFs can be assumed to be approximately rotationally symmetric with respect to the interaural axis, by definition the y-axis (see *figure 3.2.1(b)*).

Taking a closer look at the definition of the spherical harmonics in *equation 3.1.1*, we can find

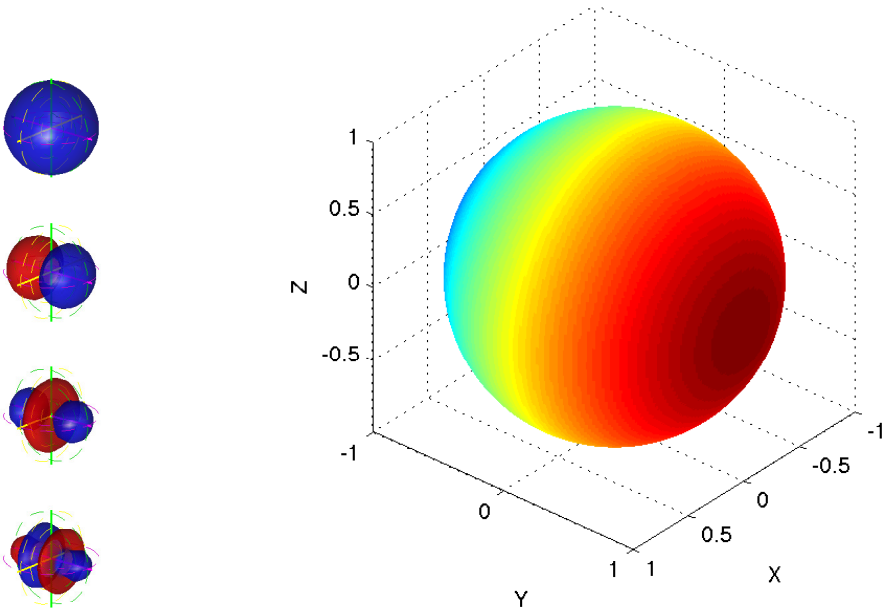
corresponding symmetry. If $m = 0$ the equation simplifies to

$$Y_n^0(\beta, \alpha) = \sqrt{\frac{(2n+1)}{4\pi}} P_n(\cos\alpha),$$

where the only dependency is $P_n = P_n^0$, the *Legendre polynomial* which represents the special case where *"no variation in the azimuthal direction"* exists [Wil99]. The angle that is normally defined as the azimuth represents the polar plane angle β here. All spherical harmonics Y_n^0 of arbitrary order $n \in [0, 1, ..., \mathrm{N}]$ and degree $m = 0$ can therefore be written as

$$Y_n^0(\beta, \alpha) = Y_n^0(\alpha),$$

since they depend solely on the lateral angle α. The spatial property resulting from that is *rotational symmetry with respect to the interaural axis* (y-axis, see *figure 3.2.1(a)*). If rotational symmetry can be assumed for the unwrapped phase, it should be representable using only spherical harmonics of degree $m = 0$.



(a) $Y_{0:3}^0$ (b) Expected unwrapped phase distribution

Figure 3.2.1: Rotational symmetry of rotated spherical harmonics and assumed spatial phase distribution of HRTFs

3.3 Discrete Spherical Harmonic Transform

Similar to [Zot09], the mathematical nomenclature for the decomposition of a set of spatially discrete values into spherical harmonics is defined as follows. For the sake of clarity, the discrete angular positions given by polar plane angle β and lateral angle α are summarized and expressed as $\boldsymbol{\theta}$. P denotes the number of known discrete positions.

Spherical harmonic expansion. Given a discrete spherical pattern $\boldsymbol{g} = [g(\boldsymbol{\theta}_1),\ ...,\ g(\boldsymbol{\theta}_\mathrm{P})]^\mathrm{T}$, its spherical harmonic expansion $\boldsymbol{g}_\mathrm{N} = [g_\mathrm{N}(\boldsymbol{\theta}_1),\ ...,\ g_\mathrm{N}(\boldsymbol{\theta}_\mathrm{P})]^\mathrm{T}$, band-limited by $n \leq \mathrm{N}$, can be written as

$$\boldsymbol{g}_\mathrm{N} = \boldsymbol{Y}_\mathrm{N}\ \boldsymbol{\gamma}_\mathrm{N}, \tag{3.3.1}$$

where

$$\boldsymbol{Y}_\mathrm{N} = \begin{bmatrix} \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_1) \\ . \\ . \\ . \\ \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_\mathrm{P}) \end{bmatrix}, \qquad \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}) = [Y_0^0(\boldsymbol{\theta}),\ ...,\ Y_\mathrm{N}^\mathrm{N}(\boldsymbol{\theta})]$$

is a matrix containing the $(\mathrm{N}+1)^2$ spherical harmonics, evaluated at the positions $\boldsymbol{\theta}_1,\ ...,\ \boldsymbol{\theta}_\mathrm{P}$ and

$$\boldsymbol{\gamma}_\mathrm{N} = \begin{bmatrix} \gamma_0^0 & \gamma_1^{-1} & \gamma_1^0 & \gamma_1^1 & ..... & \gamma_\mathrm{N}^\mathrm{N} \end{bmatrix}^\mathrm{T}$$

is a vector containing the corresponding spherical harmonic coefficients. As can be seen from equation 3.3.1, the coefficient vector $\boldsymbol{\gamma}_\mathrm{N}$ needs to be found in order to calculate the spatially continuous spherical pattern $\boldsymbol{g}_\mathrm{N}$ from the spatially discrete pattern $\boldsymbol{g}$.

Assuming that $\boldsymbol{g} = \boldsymbol{g}_\mathrm{N}$, equation 3.3.1 can be re-written as

$$\boldsymbol{g} = \boldsymbol{Y}_\mathrm{N}\ \boldsymbol{\gamma}_\mathrm{N}$$

and the inversion writes as

$$\boldsymbol{\gamma}_\mathrm{N} = \boldsymbol{Y}_\mathrm{N}^{-1}\ \boldsymbol{g}, \tag{3.3.2}$$

yielding the spherical harmonic coefficients.

3.4 Reconstruction and Interpolation

The identification of the spherical harmonic coefficients $\boldsymbol{\gamma}_\mathrm{N}$ allows the reconstruction of the spherical pattern $\boldsymbol{g}$. Depending on the complexity of $\boldsymbol{g}$, the distribution of the sampling positions $\boldsymbol{\theta}$, their number P and the maximal order N, the *reconstruction*

$$\boldsymbol{g}_\mathrm{N} = \boldsymbol{Y}_\mathrm{N}\ \boldsymbol{\gamma}_\mathrm{N} \tag{3.4.1}$$

might differ from the original pattern $\boldsymbol{g}$ (except for *hyperinterpolation*, see [Zot09]), which is due to matrix inversion errors (see chapter 3.5).

Furthermore, *interpolation* of the discrete spherical pattern $\boldsymbol{g}$ can be achieved, taking advantage of the spherical harmonics' spatial continuity. Note that the vector $\boldsymbol{\gamma}_\mathbf{N}$ does not depend on the position variable $\boldsymbol{\theta}$. The spherical harmonic expansion can therefore be evaluated at any arbitrary position $\boldsymbol{\theta}_\mathrm{des}$ on the surface of the sphere

$$g_\mathrm{N}(\boldsymbol{\theta}_\mathrm{des}) = \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_\mathrm{des})\ \boldsymbol{\gamma}_\mathrm{N}, \tag{3.4.2}$$

yielding the interpolated version of $\boldsymbol{g}$ at the desired position $\boldsymbol{\theta}_\mathrm{des}$.

3.5 Invertibility and Regularization

The matrix $\boldsymbol{Y}_\mathrm{N}$ is typically not square, unless $(\mathrm{N}+1)^2 = \mathrm{P}$, thus in general an under- or overdetermined system of equations needs to be solved. Further, $\boldsymbol{Y}_\mathrm{N}$ might be ill-conditioned depending on the regularity of the distribution of sampling positions $\boldsymbol{\theta}$ and the maximal order N. If, for example, the sampling positions are only defined on a partial sphere, as e.g. sampling distributions with a polar gap, which are quite usual for HRTFs, some sort of regularization has to be applied.

A very common method for matrix regularization is the *singular-value decomposition (SVD)* [NZK11]

$$\boldsymbol{Y}_\mathrm{N} = \boldsymbol{U}\ \boldsymbol{S}\ \boldsymbol{V}^\mathrm{T},$$

decomposing $\boldsymbol{Y}_\mathrm{N}$ into a diagonal matrix $\boldsymbol{S} = diag\{\boldsymbol{s}\}$, where $\boldsymbol{s}$ are the singular values, and two orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ containing the left and right eigenvectors, respectively. Keeping only the largest singular values, the size of $\boldsymbol{S}$ can be reduced and $\boldsymbol{U}$ and $\boldsymbol{V}$ truncated accordingly. The regularized pseudoinverse can then be calculated using the *truncated SVD*

$$\boldsymbol{Y}_N^\dagger = \tilde{\boldsymbol{V}}\ \tilde{\boldsymbol{S}}^{-1}\ \tilde{\boldsymbol{U}}^\mathrm{T},$$

where $\tilde{\boldsymbol{S}}$, $\tilde{\boldsymbol{U}}$ and $\tilde{\boldsymbol{V}}$ are the truncated versions of the original matrices.

If a well-distributed sampling grid can be assumed (i.e. no large areas without any sampling positions), another way to calculate the pseudoinverse is a **least-squares approach**:
Based on the mean square error

$$J(\boldsymbol{\gamma}_{\mathrm{N}}) = \|\boldsymbol{g} - \boldsymbol{Y}_{\mathrm{N}}\ \boldsymbol{\gamma}_{\mathrm{N}}\|^2,$$

the minimal error results from its first derivative

$$\frac{\partial J(\boldsymbol{\gamma}_{\mathrm{N}})}{\partial \boldsymbol{\gamma}_{\mathrm{N}}} = -2\left[\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{g} - \boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{N}}\boldsymbol{\gamma}_{\mathrm{N}}\right] = 0,$$

yielding

$$\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{N}}\boldsymbol{\gamma}_{\mathrm{N}} = \boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{g} \qquad \Longrightarrow \qquad \boldsymbol{\gamma}_{\mathrm{N}} = [\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{N}}]^{-1}\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\ \boldsymbol{g}.$$

The resulting pseudoinverse $\boldsymbol{Y}_{\mathrm{N}}^{\dagger} = [\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{N}}]^{-1}\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}$ inverts the spherical harmonic expansion from the left.
The least-squares error distribution depends on the distribution of the sampling positions. Voronoi tesselation can be applied to derive proper weighting factors in order to minimize the error due to the distribution of the sampling positions. Introducing a weight vector $\boldsymbol{w}$ containing the weighting factors, the *weighted pseudoinverse* can be defined as

$$\boldsymbol{Y}_{\mathrm{N}}^{\dagger} = [\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}diag\{\boldsymbol{w}\}\boldsymbol{Y}_{\mathrm{N}}]^{-1}\boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}diag\{\boldsymbol{w}\}.$$

Using the weighted pseudoinverse, equation 3.3.2 can now be modified to

$$\boldsymbol{\gamma}_{\mathrm{N}} = \boldsymbol{Y}_{\mathrm{N}}^{\dagger}\ \boldsymbol{g}, \tag{3.5.1}$$

yielding the spherical harmonic coefficients.

3.6 Definition of Maximal Order for Spatial Sampling Grids

The spatial sampling grid of the dummy head measurements used in the following chapters is depicted in *figure 3.6.1*. It covers a full sphere with a total of 1014 points and the highest density around the horizontal plane.

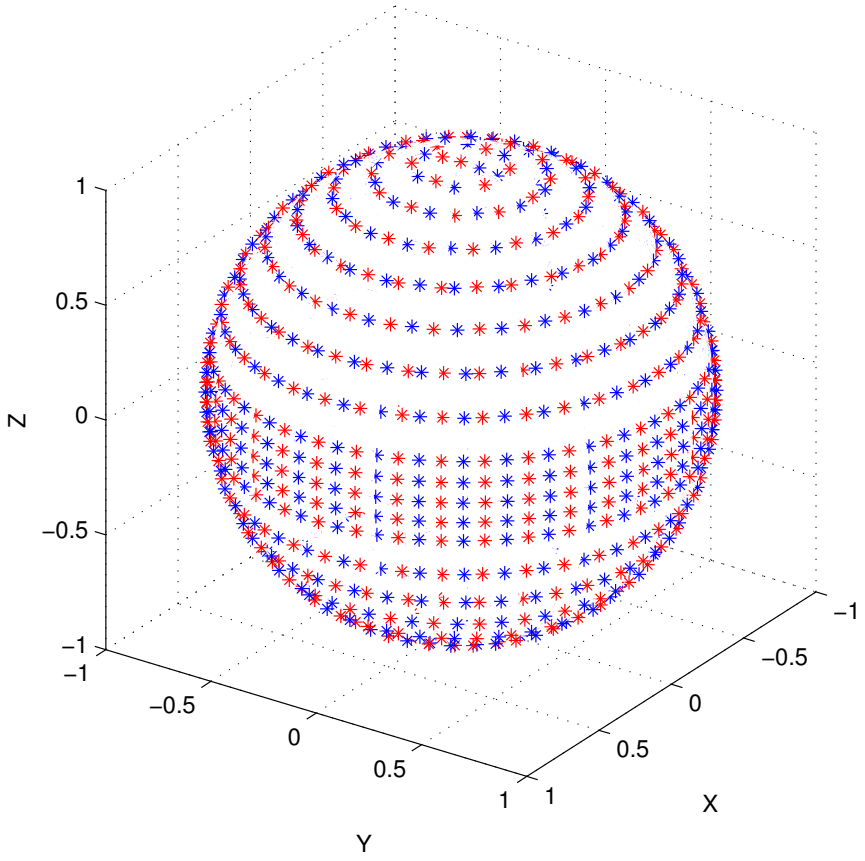


Figure 3.6.1: Spatial sampling grid and reconstruction grid $\mathbf{\Theta}_{\mathrm{rec}}$. Blue stars: Interpolation grid $\mathbf{\Theta}_{\mathrm{int}}$, red stars: Reference grid for Interpolation $\mathbf{\Theta}_{\mathrm{int,ref}}$

Reconstruction Grid. In order to recompose HRTFs as described in *equation 3.4.1*, all measured positions $\mathbf{\Theta}_{\mathrm{rec}}$ are used.
Taking the invertibility of the spherical harmonics matrix as a requirement, the maximal spherical harmonic order $\mathrm{N}_{\mathrm{rec}}$ can be defined according to a realistic condition number of $\boldsymbol{Y}_{\mathrm{N}}$ as a function of N. The maximal spherical harmonics order was set to $\mathrm{N}_{\mathrm{rec}} = 18$, resulting in a condition number $\mathrm{cond}(\boldsymbol{Y}_{\mathrm{N}_{\mathrm{rec}}}) = 3.5$ (see *figure 3.6.2*).
In order to calculate the spherical harmonic coefficients $\boldsymbol{\gamma}_{\mathrm{N}_{\mathrm{rec}}}$, a weighted pseudoinverse $\boldsymbol{Y}^{\dagger}_{\mathrm{N}_{\mathrm{rec}}}$ needs to be defined according to *equation 3.5.1*. Voronoi tesselation was applied to calculate the weighting coefficients using a Matlab toolbox provided by IRCAM.

Interpolation grid. In order to interpolate the HRTFs as described in *equation 3.4.2*, only each second position of $\mathbf{\Theta}_{\mathrm{rec}}$ is used for the decomposition: $\mathbf{\Theta}_{\mathrm{int}}$ (see *figure 3.6.1*). The remaining positions $\mathbf{\Theta}_{\mathrm{int,ref}}$ are used as a reference for the interpolated filters. Again, the maximal spherical harmonics order $\mathrm{N}_{\mathrm{int}}$ was chosen after evaluating the condition number of $\boldsymbol{Y}_{\mathrm{N}}$ as a function of N. The maximal spherical harmonics order was set to $\mathrm{N}_{\mathrm{int}} = 17$, resulting

in a condition number $\mathrm{cond}(\boldsymbol{Y}_{\mathrm{N_{int}}}) = 2.5$ (see *figure 3.6.3*). The pseudoinverse was calculated as described before.

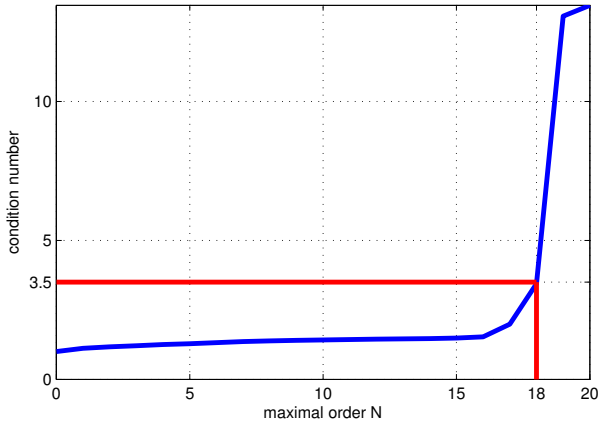


Figure 3.6.2: Reconstruction grid $\boldsymbol{\Theta}_{\mathrm{rec}}$: Condition number of $\boldsymbol{Y}_{\mathrm{N}}$

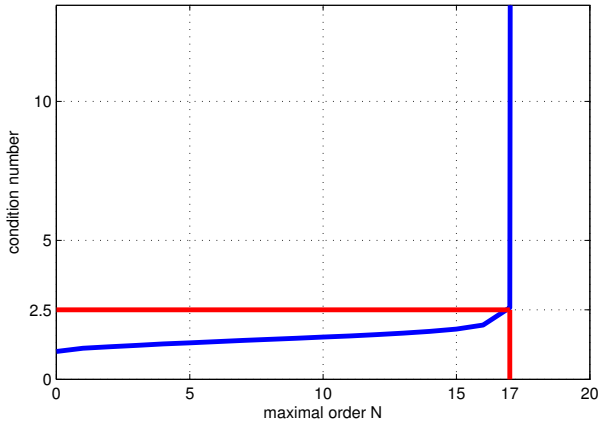


Figure 3.6.3: Interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$: Condition number of $\boldsymbol{Y}_{\mathrm{N}}$

3.7 Aliasing

In this section, the consequences of the sampling grid concerning aliasing are compared for the common and the rotated version of the DSHT.

In [Zot09], an aliasing map has been used. Consider an ideal spherical harmonic representation $\boldsymbol{g} = \boldsymbol{Y}\boldsymbol{\gamma}$, without band limitation ($\mathrm{N} = \infty$), and its band-limited counterpart $\boldsymbol{g}_{\mathrm{N}} = \boldsymbol{Y}_{\mathrm{N}}\boldsymbol{\gamma}_{\mathrm{N}}$ with $\mathrm{N} < \infty$. Ideally, all higher-order harmonics $n > \mathrm{N}$ should be suppressed in the band-limited case

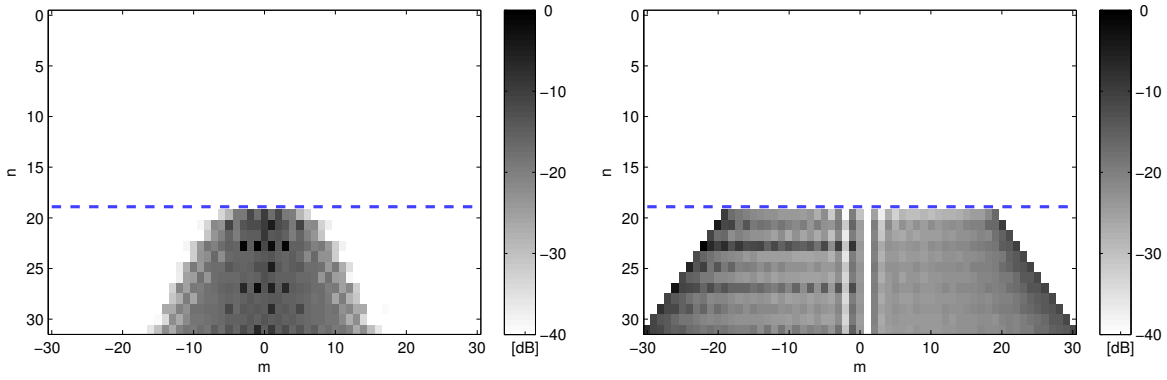
$$\begin{aligned} \boldsymbol{g}_{\mathrm{N}} &= \boldsymbol{g}, \\ \boldsymbol{Y}_{\mathrm{N}}\boldsymbol{\gamma}_{\mathrm{N}} &= \boldsymbol{Y}\boldsymbol{\gamma}, \\ \Rightarrow \boldsymbol{\gamma}_{\mathrm{N}} &= \boldsymbol{Y}_{\mathrm{N}}^{\dagger}\boldsymbol{Y}\boldsymbol{\gamma}, \end{aligned} \tag{3.7.1}$$

such that only the truncated coefficients $\boldsymbol{\gamma}_{\mathrm{N}}$ remain, i.e. $\boldsymbol{Y}_{\mathrm{N}}^{\dagger}\boldsymbol{Y} = \boldsymbol{I}$. An error measure can hence be introduced as

$$\|\boldsymbol{e}_{\boldsymbol{\gamma}}\|^2 = \| \left(\boldsymbol{Y}_{\mathrm{N}}^{\dagger}\boldsymbol{Y} - \boldsymbol{I} \right) \boldsymbol{\gamma} \|^2. \tag{3.7.2}$$

Note that the multiplication of the band-limited spherical harmonics matrix with its pseudoinverse yields the identity matrix, $\boldsymbol{Y}_{\mathrm{N}}^{\dagger}\boldsymbol{Y}_{\mathrm{N}} = \boldsymbol{I}$, provided that $\boldsymbol{Y}_{\mathrm{N}}$ is well-conditioned. Errors therefore only occur for orders above the band limitation, $n > \mathrm{N}$.

In *figure 3.7.1* the normalized aliasing errors for the reconstruction sampling grid $\boldsymbol{\Theta}_{\mathrm{rec}}$ are depicted up to $n = 30$ for the commonly used DSHT and the rotated DSHT. The blue line represents the band limitation by $\mathrm{N}_{\mathrm{rec}} = 18$. Depending on the orientation of the DSHT, very different aliasing maps result. Given a spatial pattern $\boldsymbol{g}$ which excites higher-order harmonics, these maps illustrate how much energy of a specific basis function will be mirrored into the band-limited range.



(a) Common spherical harmonic transform (b) Rotated spherical harmonic transform

Figure 3.7.1: Spatial aliasing map for $\boldsymbol{\Theta}_{\mathrm{rec}}$ [Zot09]

3.8 Reconstruction and Interpolation of Simulated HRTFs

In this section, an initial practical investigation concerning the suitability of different HRTF representations for the DSHT is performed. Furthermore, the significance of the rotated spherical harmonics for a rotationally symmetric HRTF-model is analyzed.

Filter simulation. The filters were calculated using a spherical head model presented by Duda et al. [DM98]. This model generates the transfer function between a distant sound source and an observation point on a rigid sphere. The position is defined by the distance between the center of the sphere and the sound source $\mathrm{r}_{\mathrm{source}}$, the sphere radius $\mathrm{r}_{\mathrm{head}}$ and the angle α between source and observation point, depicted in *figure 3.8.1*. The lateral angles

α derived from the spatial sampling grid $\mathbf{\Theta}_{\mathrm{rec}}$ depicted in *figure 3.6.1* were used, the sphere radius was set to $\mathrm{r_{head}} = 8.5\mathrm{cm}$ and the source distance to $\mathrm{r_{source}} = 1\mathrm{m}$. The observation point hence corresponds to the left ear position.

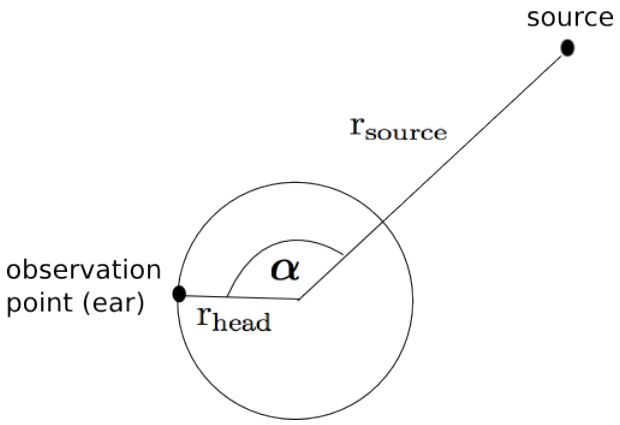


Figure 3.8.1: Spherical head model

The Filters were calculated for 129 discrete frequencies ω_l, regularly distributed between $0\mathrm{Hz}$ and $22.05\mathrm{kHz}$. This corresponds to a DFT-length $\mathrm{L} = 256$ and a sampling rate $\mathrm{f_s} = 44.1\mathrm{kHz}$. The impulse responses were then calculated via the inverse discrete fourier transform of the symmetrically complemented filter spectra. Since the filters do not depend on the polar plane angle β, they are rotationally symmetric in relation to the interaural axis.

DSHT using reconstruction grid and interpolation grid. The spherical harmonic coefficients were calculated for the impulse responses $h[\boldsymbol{\theta}, k]$, the magnitude responses $|H[\boldsymbol{\theta}, \omega_l]|$ and the unwrapped phase responses $\Phi[\boldsymbol{\theta}, \omega_l]$ using all positions $\mathbf{\Theta}_{\mathrm{rec}}$

$$\begin{aligned} \boldsymbol{\gamma}_{\mathrm{N_{rec}},\mathrm{h}}[k] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{rec}}}\ h[\mathbf{\Theta}_{\mathrm{rec}}, k], \\ \boldsymbol{\gamma}_{\mathrm{N_{rec}},|\mathrm{H}|}[\omega_l] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{rec}}}\ |H[\mathbf{\Theta}_{\mathrm{rec}}, \omega_l]|, \\ \boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi}[\omega_l] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{rec}}}\ \Phi[\mathbf{\Theta}_{\mathrm{rec}}, \omega_l], \end{aligned} \tag{3.8.1}$$

and using the subset of positions $\mathbf{\Theta}_{\mathrm{int}}$

$$\begin{aligned} \boldsymbol{\gamma}_{\mathrm{N_{int}},\mathrm{h}}[k] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{int}}}\ h[\mathbf{\Theta}_{\mathrm{int}}, k], \\ \boldsymbol{\gamma}_{\mathrm{N_{int}},|\mathrm{H}|}[\omega_l] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{int}}}\ |H[\mathbf{\Theta}_{\mathrm{int}}, \omega_l]|, \\ \boldsymbol{\gamma}_{\mathrm{N_{int}},\Phi}[\omega_l] &= \mathbf{Y}^{\dagger}_{\mathrm{N_{int}}}\ \Phi[\mathbf{\Theta}_{\mathrm{int}}, \omega_l]. \end{aligned} \tag{3.8.2}$$

Reconstruction and interpolation via Inverse DSHT. The reconstructed impulse responses $h_{\mathrm{N_{rec}}}[\boldsymbol{\theta}, k]$, magnitude responses $|H_{\mathrm{N_{rec}}}[\boldsymbol{\theta}, \omega_l]|$ and phase responses $\Phi_{\mathrm{N_{rec}}}[\boldsymbol{\theta}, \omega_l]$ were

calculated for all positions $\boldsymbol{\Theta}_{\mathrm{rec}}$ as in *equation 3.4.1*. The interpolated impulse responses $h_{\mathrm{N_{int}}}[\boldsymbol{\theta}, k]$, magnitude responses $|H_{\mathrm{N_{int}}}[\boldsymbol{\theta}, \omega_l]|$ and phase responses $\Phi_{\mathrm{N_{int}}}[\boldsymbol{\theta}, \omega_l]$ were calculated for the interpolation reference positions $\boldsymbol{\Theta}_{\mathrm{int,ref}}$ as in *equation 3.4.2*. The reconstructed magnitude and phase responses were then transformed to impulse responses via the inverse DFT of the complex spectra $H_{\mathrm{N_{rec}}}[\boldsymbol{\theta}, \omega_l] = |H_{\mathrm{N_{rec}}}[\boldsymbol{\theta}, \omega_l]|\ \mathrm{e}^{\mathrm{j}\,\Phi_{\mathrm{rec}}[\boldsymbol{\theta},\omega_l]}$ and $H_{\mathrm{N_{int}}}[\boldsymbol{\theta}, \omega_l] = |H_{\mathrm{N_{int}}}[\boldsymbol{\theta}, \omega_l]|\ \mathrm{e}^{\mathrm{j}\,\Phi_{\mathrm{int}}[\boldsymbol{\theta},\omega_l]}$.

3.8.1 Reconstruction and Interpolation Quality

Impulse response envelope error. In order to evaluate the resulting impulse responses, a simple error measure for the overall similarity of impulse responses is used. To avoid an overestimation of the impulse responses' fine structure, the envelopes of the reference impulse response h_{ref} and the manipulated impulse response h are compared. The envelope of a real signal can be calculated using the absolute value of its Hilbert transform $h_{\mathrm{env}}[k] = |\mathcal{H}\{h[k]\}|$ (*figure 3.8.2*). The impulse response envelope error is defined as

$$E_{\mathrm{IR}}(\boldsymbol{\theta}) = \frac{1}{\mathrm{K}} \sum_{k=0}^{\mathrm{K}-1} \frac{|\ |\mathcal{H}\{h_{\mathrm{ref}}[\boldsymbol{\theta}, k]\}| - |\mathcal{H}\{h[\boldsymbol{\theta}, k]\}|\ |}{max(|\mathcal{H}\{h_{\mathrm{ref}}[\boldsymbol{\theta}, k]\}|)}.$$

Since this error measure is only used for comparison, it is depicted in dB with 0dB corresponding to the maximal occurring value.

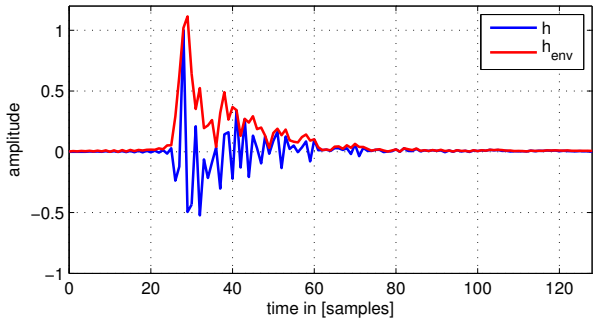


Figure 3.8.2: Impulse response h and envelope h_{env}

Mean errors and examples. In *figure 3.8.3* the mean of the impulse response envelope error E_{IR} over all positions is depicted. It can be ascertained that the filter representation using magnitude and unwrapped phase is much more adequate than the time domain representation concerning the requirements of the DSHT. Unsurprisingly, the interpolation quality is generally worse than the reconstruction quality.

The impulse responses corresponding to the worst-cases in terms of the error can be compared in *figure 3.8.4*. Independent of the filter representation, the largest errors occur at contralateral positions. When using the DSHT of the impulse responses, obvious 'pre-delays' can be ob-

served as artifacts. They apparently result from crossfading the peaks of all impulse responses. The highest peaks can be found in impulse responses at ipsilateral positions, which have a shorter onset delay than impulse responses at the extreme contralateral positions where the error is maximal. Using the DSHT of magnitude and unwrapped phase, even these contralateral impulse responses can be reconstructed and interpolated quite accurately. The worst-case is a widened peak which does not reach the full height while the temporal structure of the impulse response remains virtually unchanged. Although one might argue that a low reconstruction and interpolation accuracy for contralateral positions might not be perceptually relevant, the 'pre-delays' might have disastrous effects on the localization cues, especially on the ITD.

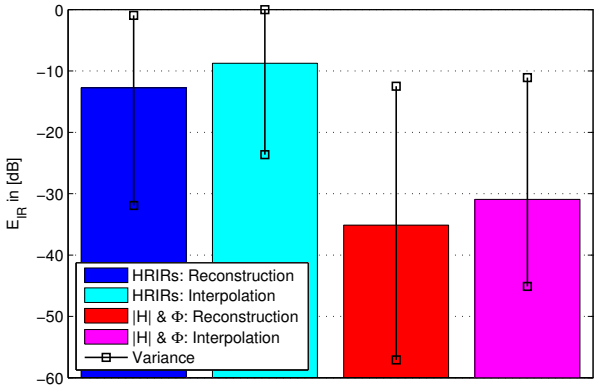
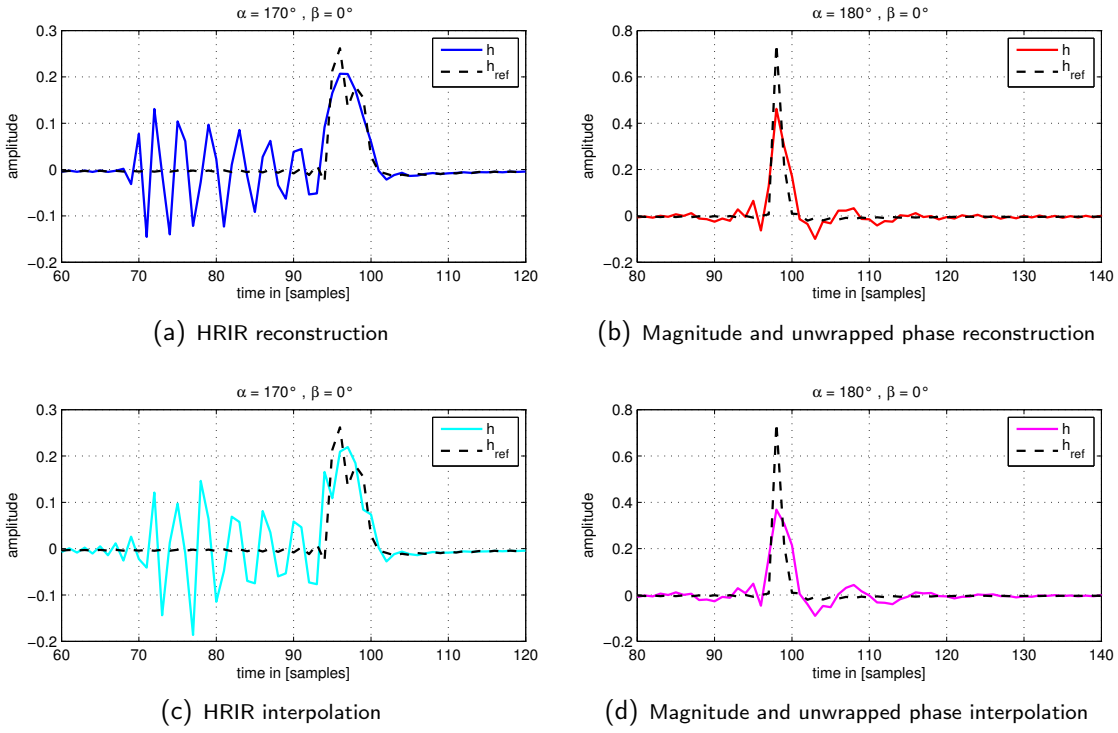


Figure 3.8.3: Mean values of E_{IR}



(a) HRIR reconstruction

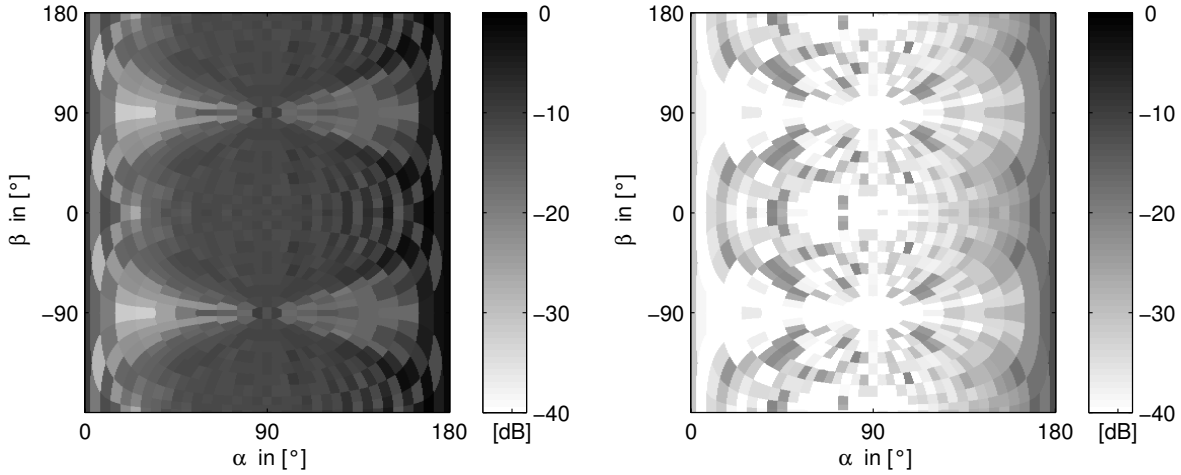
(b) Magnitude and unwrapped phase reconstruction

(c) HRIR interpolation

(d) Magnitude and unwrapped phase interpolation

Figure 3.8.4: Worst-case impulse responses

Spatial error distribution. The spatial distribution of the impulse response envelope error E_{IR}, depicted in *figure 3.8.5*, is largely similar for reconstruction and interpolation. The overall error level is much higher when using direct DSHT of the impulse responses, whereas the separate DSHTs of magnitude and unwrapped phase seem to achieve much more accurate results. The error tends to increase from the ipsilateral to the contralateral side of the sphere.



(a) HRIR reconstruction

(b) Magnitude and unwrapped phase reconstruction

(c) HRIR interpolation

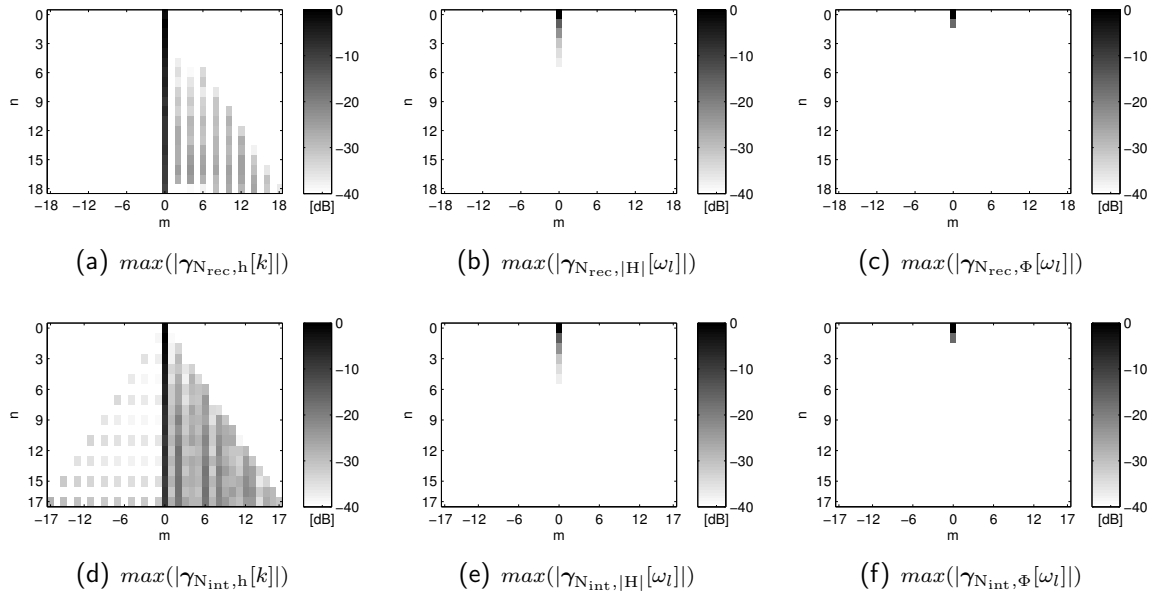
(d) Magnitude and unwrapped phase interpolation

Figure 3.8.5: Spatial distribution of impulse response envelope error E_{IR}

3.8.2 Most Important Basis Functions

Calculation of maximal coefficients. Since the spherical harmonic transform is applied to each frequency bin or each sample separately, the calculation of the overall importance of a specific coefficient over all frequencies or all samples requires some consideration. For the coefficients derived from magnitude ($\boldsymbol{\gamma}_{\mathrm{N_{rec},|H|}}[\omega_l]$ and $\boldsymbol{\gamma}_{\mathrm{N_{int},|H|}}[\omega_l]$) and unwrapped phase ($\boldsymbol{\gamma}_{\mathrm{N_{rec},\Phi}}[\omega_l]$ and $\boldsymbol{\gamma}_{\mathrm{N_{int},\Phi}}[\omega_l]$), the absolute values were normalized according to their maximum

value per frequency bin and weighted with $1/f$ for frequencies greater than $500\mathrm{Hz}$ to avoid an overestimation of high-frequency components. For the coefficients derived from the impulse responses ($\boldsymbol{\gamma}_{\mathrm{N_{rec}},\mathrm{h}}[k]$ and $\boldsymbol{\gamma}_{\mathrm{N_{int}},\mathrm{h}}[k]$), the coefficients were filtered according to the $1/f$-weighting, afterwards the absolute values were normalized according to their overall maximum per sample and position, thus emphasizing the importance of high-amplitude components. The maxima of the weighted coefficients over time/frequency are depicted in *figure 3.8.6*.



(a) $max(|\boldsymbol{\gamma}_{\mathrm{N_{rec}},\mathrm{h}}[k]|)$ (b) $max(|\boldsymbol{\gamma}_{\mathrm{N_{rec}},|\mathrm{H}|}[\omega_l]|)$ (c) $max(|\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi}[\omega_l]|)$

(d) $max(|\boldsymbol{\gamma}_{\mathrm{N_{int}},\mathrm{h}}[k]|)$ (e) $max(|\boldsymbol{\gamma}_{\mathrm{N_{int}},|\mathrm{H}|}[\omega_l]|)$ (f) $max(|\boldsymbol{\gamma}_{\mathrm{N_{int}},\Phi}[\omega_l]|)$

Figure 3.8.6: Maximal weighted spherical harmonic coefficients in [dB], derived from HRIRs (left), magnitude (middle) and unwrapped phase (right) for reconstruction grid $\boldsymbol{\Theta}_{\mathrm{rec}}$ (top) and interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$ (bottom).

Interpretation. The consequences of the rotation of the spherical harmonics described in chapter 3.1 can be observed in *figure 3.8.6*. The simulated filters used here are rotationally symmetric in relation to the interaural axis. Therefore the most important coefficients are clearly those with degree $m = 0$, corresponding to the rotationally symmetric rotated spherical harmonics (see chapter 3.2). The coefficients with $m \neq 0$ derived from magnitude and unwrapped phase have much smaller values. Coefficients derived from the impulse responses have their largest values for $m = 0$, while significant values for other degrees apparently exist. The values of coefficients with $m \neq 0$ increase for the interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$, i.e. using less sampling positions. Obviously, the fine structure of the impulse responses results in a higher sensitivity to spatial aliasing (see chapter 3.7). The spatial sampling grid hence seems to be insufficient for the DSHT of impulse responses while it seems to suffice for the DSHTs of magnitude and unwrapped phase.

Chapter 4

Phase Unwrapping

4.1 Spatial Phase Distribution

Measured dummy head HRTFs. For all further investigations, a set of dummy head HRTFs will be used which has been measured at the IRCAM. A HEAD acoustics HSU III mannequin was used, which provides a relatively simple reproduction of the human head, torso and ears.

Figure 4.1.1: HEAD acoustics HSU III

Unlike usual data sets, measured responses of a dummy can be determined over the whole sphere without polar gap below. The sampling grid is described in chapter 3.6 and depicted in *figure 3.6.1*. Note that this data set represents realistic HRTFs in terms of their spatial complexity, as opposed to the spherical head model simulation used in chapter 3.8.

Spatial phase continuity. The application of DSHT to the HRTFs' phase implicitly requires a continuous phase distribution which is assumed to be defined by the unwrapped phase. This can be based on the constraint that the spherical harmonic transform demands a *"square integrable function on a sphere"* [Wil99], i.e. a spherical pattern without discontinuities. It is

also obvious when considering a local approach for phase interpolation because the unwrapped phase has to be consistent with the true group delay to produce a meaningful result (see chapter 2.3.2).

Note that the HRTFs are only available for a set of discrete sampling positions and hence need to be sampled dense enough to provide correct interpolation. Spatial phase continuity can only be granted if the unwrapped phase differences between adjacent sampling positions do not exceed π. The directly accessible phase information contained in the measured responses is the wrapped phase, a discontinuous function of frequency, which may be discontinuously distributed for each discrete frequency ω_l in space in case of spectral notches or corruption by noise. The conversion from wrapped to unwrapped phase, which has so far been applied in a spectral sense (as described in chapter 2.3.2), hence deserves some further notice.

Visual comparison. In order to clarify the spatial impacts of the spectral phase unwrapping, the wrapped phase and spectrally unwrapped phase distributions in space are compared here for specific discrete frequencies. The previously assumed rotationally symmetric properties of the unwrapped phase distribution appear as vertical lines in the following plots, discontinuities as sudden color changes.

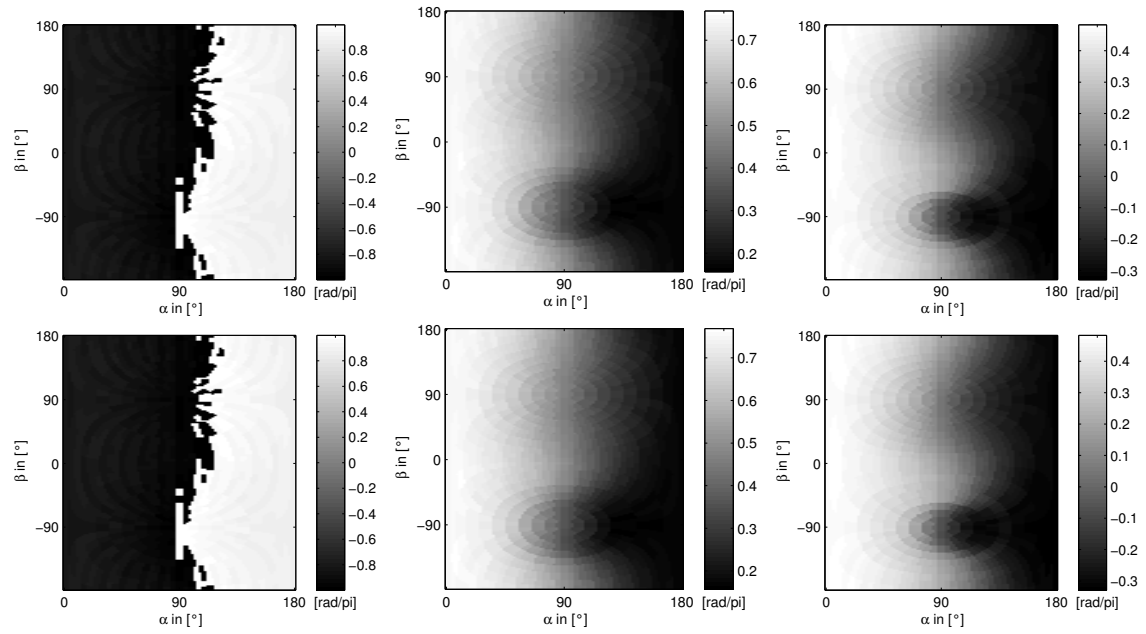


Figure 4.1.2: Top: wrapped phase in $[\frac{rad}{\pi}]$, bottom: spectrally unwrapped phase in $[\frac{rad}{\pi}]$ for frequencies 170Hz, 350Hz and 500Hz, from left to right

The wrapped and unwrapped phase distributions for $f = 170\text{Hz}$ are depicted in *figure 4.1.2* on the left. Considering the corresponding wavelength of $\lambda \approx 2m$, the absolute phase differences must be very small. All values of the unwrapped phase distribution lie inside the interval $]-\pi, \pi]$, indicating that no spectral unwrapping was necessary up to $f = 170\text{Hz}$. Still,

the change from $-\pi$ to π indicates spatial discontinuity of the spectrally unwrapped phase distribution for that frequency bin, displayed as a transition from white to black in the region of $\alpha \approx 90°$. The unwrapped phase distributions for $f = 350\text{Hz}$ and $f = 500\text{Hz}$ (middle and right in *figure 4.1.2*) also equal the corresponding wrapped phase distributions, which do not show any spatial discontinuities. As for the assumption of rotational symmetry, it can be found that while the assumption seems to hold in general, there are some deviations from it, especially below the head due to the torso.

The wrapped phase distributions for $f = 850\text{Hz}$, $f = 1\text{kHz}$ and $f = 1.5\text{kHz}$ depicted in *figure 4.1.3* all show discontinuities. As opposed to the unwrapped phase distribution for $f = 170\text{Hz}$, spectral phase unwrapping neutralizes those discontinuities by adding 2π-multiples. Concerning the assumed symmetric properties of the unwrapped phase distribution, it can be said that while a global rotational symmetry can be observed, the deviations from it seem to increase with frequency, again with maximal deviations located below the head.

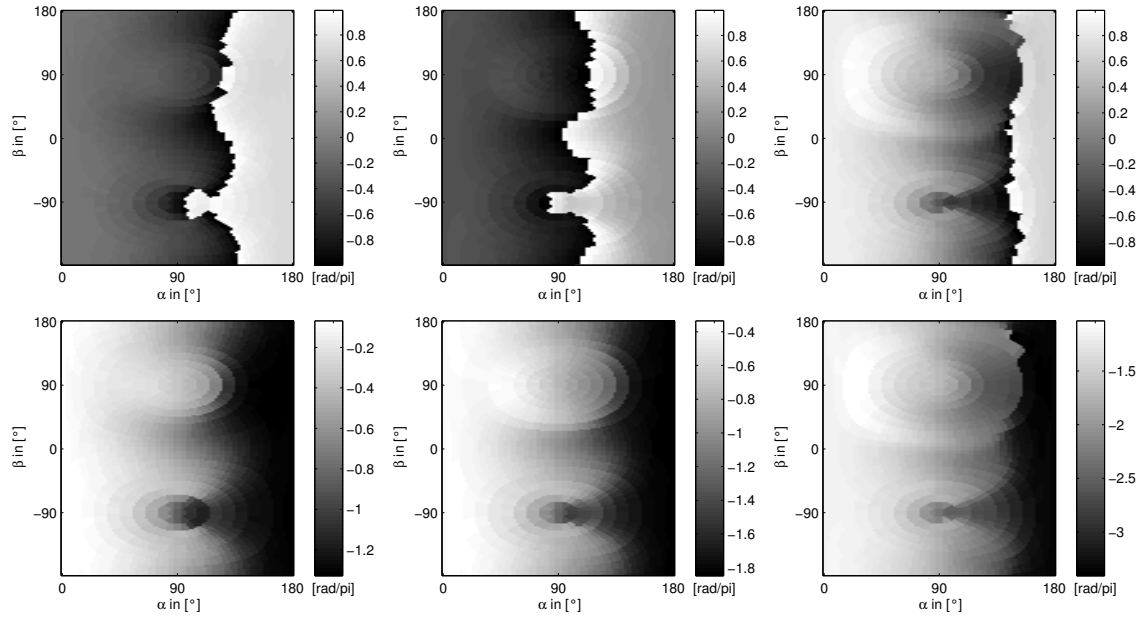


Figure 4.1.3: Top: wrapped phase, bottom: unwrapped phase for frequencies 850Hz, 1kHz, 1.5kHz, from left to right

For the relatively high frequencies depicted in *figure 4.1.4*, the wrapped phase distributions become much more complex due to very small wavelengths. While the unwrapped phase distribution for $f = 3\text{kHz}$ seems to be largely spatially continuous, for $f = 8\text{kHz}$ discontinuity becomes obvious ($\alpha = 120°,\ \beta = -80°$). Taking a closer look at the colour scaling, even more discontinuities can be identified, mainly on the contralateral side ($\alpha > 90°$). The unwrapped phase distribution for $f = 10\text{kHz}$ shows a whole area with discontinuities around its borders, while the significant discontinuity observed for 8kHz remains, indicating a propagation of the discontinuities over frequency. A slight rotational symmetry of the unwrapped phase distribution can still be observed, but the patterns become more complex with increasing

frequency. Due to shadowing effects for small wavelengths, the largest deviations can be found in the region of the torso and generally on the contralateral side.

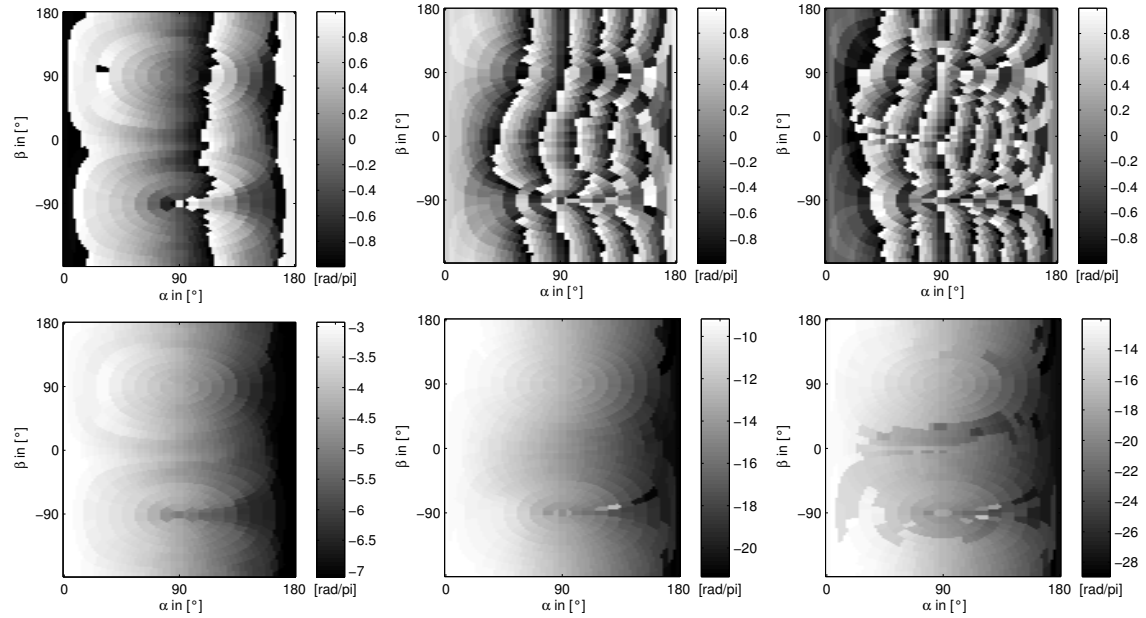


Figure 4.1.4: Top: wrapped phase, bottom: unwrapped phase for frequencies $3\mathrm{kHz}$, $8\mathrm{kHz}$, $10\mathrm{kHz}$, from left to right

Conclusions. Concluding this investigation, it can be said that *spectral phase unwrapping does not necessarily produce a continuous phase distribution in space*. Since the aim is to find a *spatially* continuous phase distribution for each discrete frequency bin, the *spatial* relation between phase values of adjacent points in space should be the criterion for phase unwrapping. Using spectral phase unwrapping, the *spectral* relation between phase values of adjacent frequency bins is used as a criterion for unwrapping.

4.2 The Concept of Phase Unwrapping

Phase unwrapping in audio applications is usually performed spectrally, i.e. from lowest to highest frequency bin. In scientific fields such as image processing and geodesy, the unwrapped phase is defined in a spatial sense, applying the unwrapping procedure to adjacent points in space. The basic unwrapping operation is usually defined as described in the following [Mat06, Spa95] and can be applied spectrally or spatially.

Any complex number $Z = a + j \cdot b$ can be expressed as

$$Z = |Z|\ e^{j\phi} \qquad \text{with } \phi = \arctan\left(\frac{b}{a}\right),\ \phi\ \in\ ]-\pi, \pi]$$

The phase angle or wrapped phase ϕ does not necessarily equal the absolute phase Φ, because

its complex representation is 2π-periodic. The phase information is restricted to the interval $]-\pi,\pi]$, neglecting any additional 2π-multiples, which can be written as the phase wrapping operation from *equation 2.3.2*

$$\begin{aligned}\phi &= \mathcal{W}(\Phi) \\ &= \mathrm{mod}_{2\pi}(\Phi+\pi)-\pi.\end{aligned}$$

The recovery of the neglected 2π-multiples using sequences of complex numbers is called phase unwrapping.
Considering a sequence $\Phi[m]$ of M adjacent absolute phase values, neighbouring in a spatial or spectral sense, we can define the linear differences by

$$\Delta\Phi[m] = \Phi[m] - \Phi[m-1]. \tag{4.2.1}$$

From *equation 4.2.1* comes immediately

$$\sum_{m=1}^{\mathrm{M}} \Delta\Phi[m] = \Phi[\mathrm{M}] - \Phi[0], \tag{4.2.2}$$

because the other values cancel out each other. Introducing the so-called *Itoh-Condition*

$$|\Delta\Phi[m]| \leq \pi, \quad \forall m \tag{4.2.3}$$

which can be understood as a sampling theorem, it can now be assumed

$$\mathcal{W}(\Delta\phi) = \Delta\Phi, \tag{4.2.4}$$

since the wrapped differences between the adjacent wrapped phase values ϕ must also lie in the interval $]-\pi,\pi]$. With this relation, *equation 4.2.2* can be re-written as

$$\Phi[\mathrm{M}] = \sum_{m=1}^{\mathrm{M}} \mathcal{W}(\Delta\phi[m]) + \Phi[0]. \tag{4.2.5}$$

Each value of the absolute phase Φ can hence be calculated from the wrapped phase ϕ, provided there is a starting value $\Phi[0]$.

This relation holds only if 1) the Itoh-Condition is fulfilled and 2) the available wrapped phase values are ideal, i.e. $\phi = \mathcal{W}(\Phi)$. Due to corruption by noise and other errors, measured data might not satisfy the second condition, while the compliance with the Itoh-Condition depends on the spatial sampling or the frequency resolution, respectively. Since *equation 4.2.5* can also be written recursively as

$$\Phi[\mathrm{M}] = \mathcal{W}(\Delta\phi[\mathrm{M}]) + \Phi[\mathrm{M}-1], \tag{4.2.6}$$

possible *unwrapping errors propagate along the path that the operation is applied to*. When spectral phase unwrapping is used, errors propagate over frequency. Once the spectrally unwrapped phase distribution for a certain frequency bin is incorrect, the distributions for all following bins will also be corrupted. The relevance of this effect depends mainly on the quality of the HRTF-measurement and the post-processing applied for enhancement of the measured data.

4.3 Spherical Phase Unwrapping

When spatial phase unwrapping is applied, errors propagate along the path in space that it is applied to. In the so-called *path-following* image processing approaches, the choice of the best paths is the essential problem. Although seemingly trivial in nature, phase unwrapping is a very complex problem in 2-D or 3-D. The aim is to find the one 'true' spatial pattern without discontinuities, which is heavily data-dependent. In order to gain robustness against corrupted data, a lot of methods have been published. They range from *Branch Cut*-algorithms, where 'bad' regions are identified and path limits are set accordingly, to *quality guided* algorithms, where the order of the unwrapping depends on the reliability of the available values [GZW88, Mat06, CMM$^+$02]. Other approaches are the *minimum norm* methods, which are based on error minimization from a global perspective. All those approaches have one essential drawback when it comes to spherical problems: They usually work on rectangular equidistant sampling grids, i.e. planes in 2-D or cuboids in 3-D. The phase unwrapping problem considered in this thesis, however, is spherically distributed.

4.3.1 Concept of Neighboring Points

In order to apply spatial unwrapping over a spherical sampling grid, the first step is to find meaningful relations between the sampling points in terms of their spacing.

Cylindrical projection. The surface of a sphere can be depicted as a cylindrical projection on a plane as in *figure 4.3.1*, a representation that seemingly facilitates the use of standard 2D image processing phase unwrapping algorithms. This is misleading because two of the plane's edges (top and bottom in this case) border each other in fact. If spatial phase unwrapping was applied to this plane, the spatial coherence between $\beta = 180°$ and $\beta = -180°$ would be disregarded. Furthermore, the sampling grid is not equidistant, and the true distances between the points are further distorted by the projection. In [CR99] Constantini and Rosen described a method for phase unwrapping on a non-rectangular grid, based on the assumption that due to noise only sparse data from a measurement on rectangular grid are available.

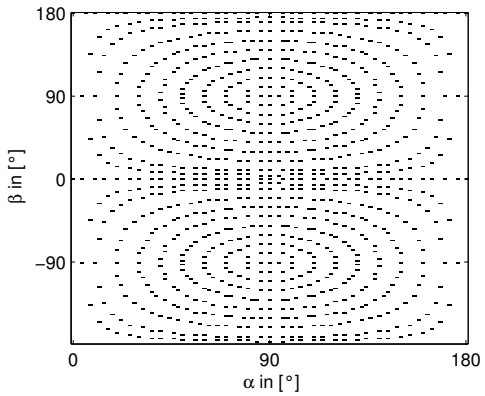


Figure 4.3.1: Sampling points as cylindrical projection

In order to identify the adjacent points for each available point in the sparse data, they utilized the concept of neighbouring points, which has been introduced by Boris Delaunay [Del34] in 1934.

Delaunay triangulation. The two-dimensional *Delaunay triangulation* finds the triangles of a set of points such that none of those points lies inside the circumcircle of any triangle [Wik], see *figure 4.3.2*. Each point is part of several triangles, the other vertices of those triangles are its neighbours. The Delaunay triangulation can be extended to 3-D considering circumscribed spheres around the triangles.
An algorithm that finds the Delaunay triangles on the convex hull of a set of points in 3-D is described in [BDH96] and can be conveniently used as a Matlab-function ('convhulln'). In *figure 4.3.3*, the Delaunay triangulation calculated via this *quickhull algorithm for convex hulls* from Barber et al. is depicted along with an example reference point and its neighbours.
Now the spatial relations between the sampling points can be defined, resulting in a *set of neighbours for each sampling position*.

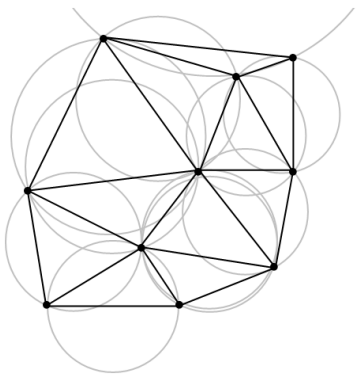


Figure 4.3.2: Principle of 2-D Delaunay triangulation [Wik]

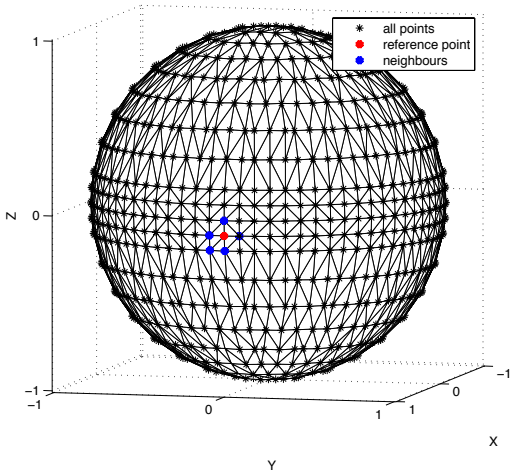
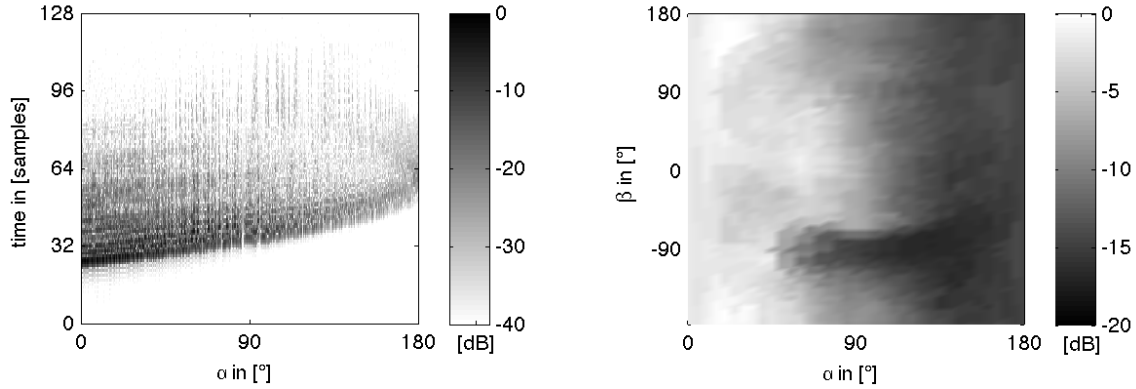


Figure 4.3.3: Delaunay triangulation of measurement grid and exemplary neighbouring points

4.3.2 Algorithm

Concept. As mentioned before, the reliability of phase unwrapping is heavily dependent on the available data along the path of unwrapping. Since unwrapping errors propagate along this path, it seems reasonable to unwrap 'good' values first and 'bad' values at last. This concept is called *quality guided phase unwrapping*. For the calculation of meaningful quality criteria several approaches can be found in literature [Mat06].
Since unwrapping errors can be mainly traced back to measurement values corrupted by noise, the signal-to-noise-ratio (SNR) is probably the best criterion. The SNR obviously depends on the level of the measured signals, hence a very simple quality measure is the energy distribution. Considering a simplified model of the human head as a sphere, thus neglecting the influence of the torso, the energy distribution of HRTFs can be assumed to decrease with the lateralisation of the sound source from ipsilateral to contralateral side. In reality this assumption does not perfectly hold because of shadowing effects produced by the torso (see *figure 4.3.4*). Still, the spherical phase unwrapping algorithm presented here is based on the constraint, that the *phase shall be unwrapped from ipsilateral to contralateral side, taking advantage of a priori knowledge about the data*.



(a) Temporal envelopes in [dB] versus lateral angle (b) Energy distribution

Figure 4.3.4: Energy distribution of left ear HRIRs

Description of the algorithm. Before starting the algorithm, the starting point $\boldsymbol{\theta}_0$ needs to be defined. According to the concept of unwrapping from ipsilateral to contralateral side, $\boldsymbol{\theta}_0$ was defined as the point closest to the interaural axis on the ipsilateral side. Φ denotes unwrapped and ϕ wrapped phase values. $\boldsymbol{\theta}_{\mathrm{ref}}$ denotes the current reference point, $\boldsymbol{\Theta}_{\mathrm{Neighbours,ref}}$ denotes the positions of the neighbours of $\boldsymbol{\theta}_{\mathrm{ref}}$. $\boldsymbol{\Theta}_{\mathrm{wrapped}}$ and $\boldsymbol{\Theta}_{\mathrm{unwrapped}}$ denote the positions of wrapped and already unwrapped phase values, respectively. The indices of the lateral angles α are used accordingly.

$\boldsymbol{\theta}_{\mathrm{ref}} \stackrel{!}{=} \boldsymbol{\theta}_0$	set: first reference point is starting point
$\Phi[\boldsymbol{\theta}_{\mathrm{ref}}] \stackrel{!}{=} \phi(\boldsymbol{\theta}_{\mathrm{ref}})$	set: Unwrapped phase of starting point is wrapped phase of starting point

Mathematical / Logical	**Verbal**
A1) Unwrap neighbours	
$\forall\ \boldsymbol{\theta}\ \in\ \boldsymbol{\Theta}_{\mathrm{Neighbours,ref}} \cap \boldsymbol{\Theta}_{\mathrm{wrapped}}$	for all wrapped neighbours of reference point:
$\Phi[\boldsymbol{\theta}] = \Phi(\boldsymbol{\theta}_{\mathrm{ref}}) + \mathcal{W}(\phi(\boldsymbol{\theta}) - \phi(\boldsymbol{\theta}_{\mathrm{ref}}))$	unwrap phase relative to reference point, see *equation 4.2.6*
A2) Smooth result: Absolute phase must decrease	
$\Phi[\boldsymbol{\theta}] > mean\{\Phi[\alpha_{\mathrm{ref}} \geq \alpha > \alpha_{\mathrm{ref}} - \Delta_\alpha]\}$ $=> \Phi[\boldsymbol{\theta}] = \Phi[\boldsymbol{\theta}] - 2\pi$	while mean of unwrapped values in the lateral region is smaller than current unwrapping result: subtract 2π
B) Find new reference point	
$\boldsymbol{\theta}_{\mathrm{ref_{new}}} \in \boldsymbol{\Theta}_{\mathrm{unwrapped}}$	*new reference point has* I. already been unwrapped
$\boldsymbol{\Theta}_{\mathrm{Neighbours,ref_{new}}} \cap \boldsymbol{\Theta}_{\mathrm{wrapped}} \neq \emptyset$	II. at least one wrapped neighbour
$\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathrm{Neighbours,ref_{new}}} \cap \boldsymbol{\Theta}_{\mathrm{unwrapped}}$ $\lvert \Phi[\boldsymbol{\theta}_{\mathrm{ref_{new}}}] - \Phi[\boldsymbol{\theta}] \rvert < \pi$	III. no discontinuities to unwrapped neighbours
$\lvert \alpha_{ref_{new}} - \alpha_0 \rvert \stackrel{!}{=} min$	IV. minimal lateral difference to starting point
C) Set new reference point, repeat	
$\boldsymbol{\theta}_{\mathrm{ref}} = \boldsymbol{\theta}_{\mathrm{ref_{new}}}$	repeat algorithm with new reference point
$\boldsymbol{\Theta}_{\mathrm{wrapped}} = \emptyset \quad => \ stop$	untill all points are unwrapped

The constraint in *B)* III. might not be fulfilled if the Itoh-Condition does not hold for higher frequencies. *B)* III. and *B)* IV. can therefore be replaced by

$$\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathrm{Neighbours,ref_{new}}} \cap \boldsymbol{\Theta}_{\mathrm{unwrapped}}$$

$$\sum |\Phi[\boldsymbol{\theta}_{\mathrm{ref_{new}}}] - \Phi[\boldsymbol{\theta}]| \stackrel{!}{=} min,$$

taking the position with the least discontinuities to its unwrapped neighbours. This has been used for frequencies above $10\mathrm{kHz}$ (see chapter 4.3.3).

Performance. The spherical phase unwrapping algorithm has been implemented as a Matlab-function called *'SphericalUnwrap'*. Since it finds the paths iteratively, it needs about $0.7\mathrm{P}$ iterations to fully unwrap a spherical phase distribution with P points. Using the 1014-point sampling grid, the average calculation time for one iteration was $\sim 7ms$. With around 700 iterations this adds up to $\sim 5s$ per frequency and hence ~ 11 minutes for all 129 frequency bins. The computation was carried out using an up-to-date personal computer with a 2.66 Ghz Intel Core i7 processor. Compared to spectral unwrapping using the 'unwrap'-function in Matlab ($\sim 80ms$ for all frequencies and positions), the algorithm is very expensive in terms of computing time.

Progression. The progression of the spherical phase unwrapping algorithm for $f = 7\mathrm{kHz}$ is depicted in *figure 4.3.5*.

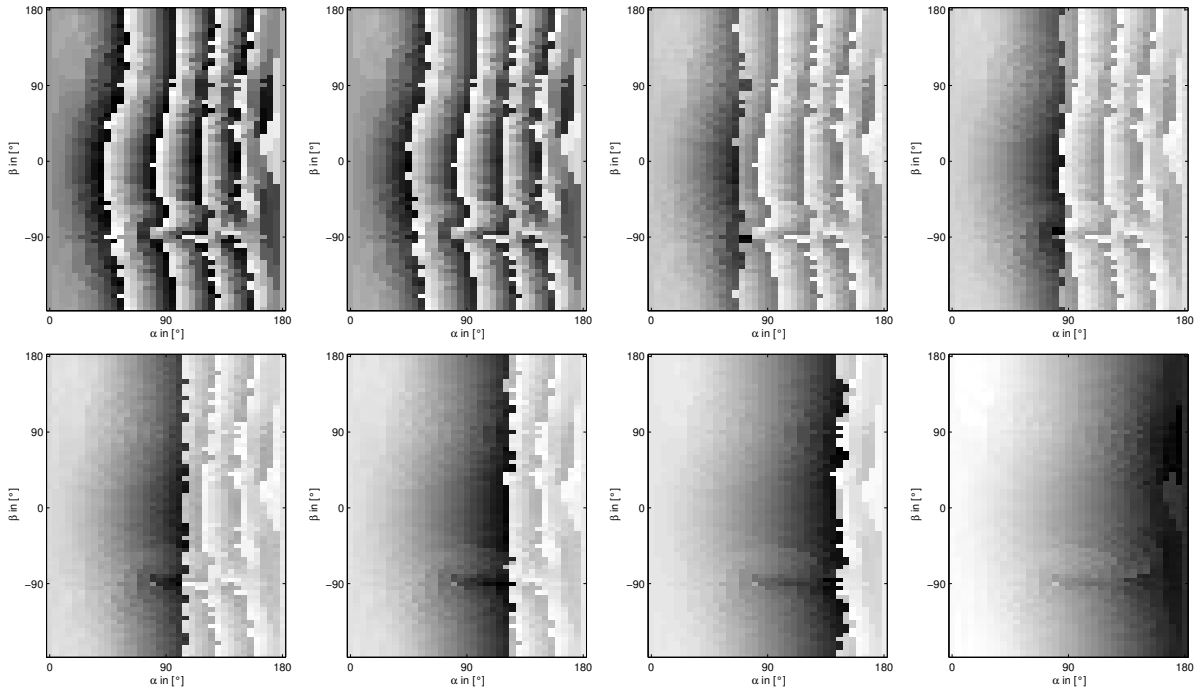


Figure 4.3.5: Progression of spherical phase unwrapping algorithm for $f = 7\mathrm{kHz}$. Top, left to right: Iterations 0, 100, 200, 300. Bottom, left to right: Iterations 400, 500, 600, 685

4.3.3 Boundaries

According to the Itoh-Condition (see *equation 4.2.3*), the reliability of the spherical phase unwrapping operation depends on the difference between spatially adjacent absolute phase values. In other words, if the absolute difference between the true absolute phase values Φ of neighbouring points exceeds π, the spacing of the sampling grid is too coarse for an unique interpretation. This is a problem of spatial aliasing. Since the true absolute phase is not available, the consistency of the absolute phase distribution and the sampling grid with the Itoh-Condition is investigated using the linear phase. The linear phase can be assumed to cover the major component of the absolute phase and can easily be calculated from the onset delay as

$$\Phi_{\mathrm{lin}}(\omega_l) = -\omega_l \ \tau.$$

The onset delay was derived from 1) the measurement data via onset extraction and 2) a spherical onset delay simulation (see annex, chapters 8.2 and 8.3). In both cases, the reconstruction sampling grid $\mathbf{\Theta}_{\mathrm{rec}}$ was used, i.e. all 1014 available points from the measurement. The maximal angular difference between neighboring points in this data set is $\sim 11^{\circ}$.

Discontinuity measure. The absolute linear phase differences between each position and its neighbours were calculated and discontinuities simply counted. The discontinuity measure $E_{\mathrm{dis}}(\boldsymbol{\theta})$ was then calculated per position and frequency by dividing the number of discontinuities by the respective number of neighbours:

$$E_{\mathrm{dis}}(\boldsymbol{\theta}) = 100 \cdot \frac{\eta_{\mathrm{dis}}(\boldsymbol{\theta})}{\eta_{\mathrm{Neighbours}}(\boldsymbol{\theta})} \quad \text{in } [\%], \tag{4.3.1}$$

where $\eta_{\mathrm{dis}}(\boldsymbol{\theta})$ denotes the number of absolute phase differences greater than π between $\boldsymbol{\theta}$ and its neighbours and $\eta_{\mathrm{Neighbours}}(\boldsymbol{\theta})$ denotes the number of the neighbours of $\boldsymbol{\theta}$.

Discontinuity measure of simulated linear phase. *Figure 4.3.6* depicts the spatial distribution of the discontinuity measure of the *simulated* linear phase for different frequencies. Since the simulated linear phase distribution is not very complex and even rotationally symmetric around the interaural axis, discontinuities appear only above $10kHz$. Clearly, the simplicity of the spherical model used for linear phase simulation provides an insufficient approximation in this context (compare *figure 4.3.7*).

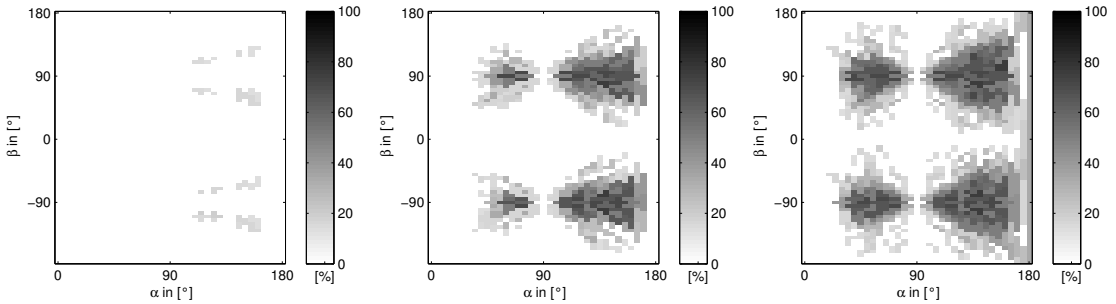


Figure 4.3.6: *Simulated linear phase*: Discontinuity measure in [%] for $f = 10\mathrm{kHz}$, $f = 15\mathrm{kHz}$, $f = 20\mathrm{kHz}$, from left to right

Discontinuity measure of extracted linear phase. In *figure 4.3.7*, the spatial distribution of the discontinuity measure of the *extracted* linear phase is depicted for different frequencies. While for $f = 3\mathrm{kHz}$ the sampling grid seems to be sufficient for all positions except the ones below the torso, the regions and quantities of the discontinuity measure increase with frequency. The regions of discontinuity tend to be restricted to the lower hemisphere for frequencies below $5\mathrm{kHz}$ and enter the upper hemisphere for frequencies above $5\mathrm{kHz}$.

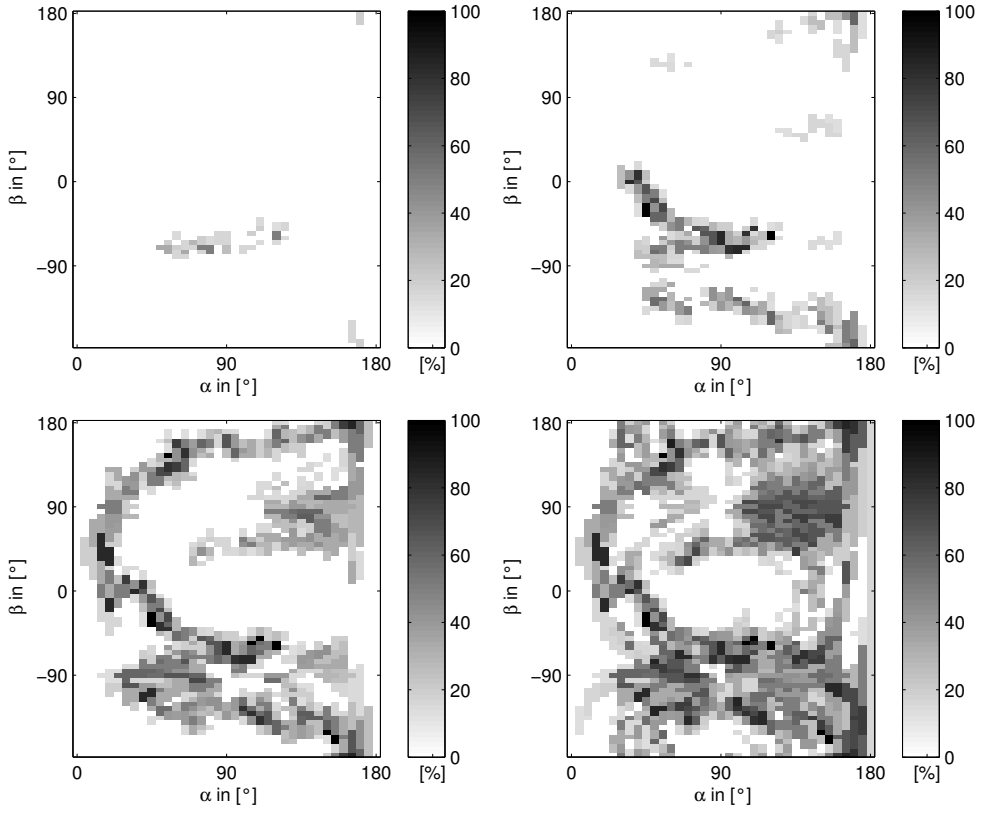


Figure 4.3.7: *Extracted linear phase*: Discontinuity measure in [%] for $f = 3\mathrm{kHz}$ (top left), $f = 5\mathrm{kHz}$ (top right), $f = 10\mathrm{kHz}$ (bottom left) and $f = 15\mathrm{kHz}$ (bottom right)

Mean discontinuity measure. The overall discontinuity measures depicted in *figure 4.3.8* show their frequency dependence. While the simulated linear phase can be accurately sampled and hence correctly unwrapped up to $10\mathrm{kHz}$, the sampling of the extracted linear phase starts to increasingly fail around $f = 3\mathrm{kHz}$, leading to a violation of the Itoh-Condition. Although this value might be slightly distorted by onset delay extraction errors, the *critical frequency for spherical phase unwrapping can still be assumed as* $\mathrm{f}_{\mathrm{Itoh}} = 3\mathrm{kHz}$. Above that frequency, the Itoh-Condition might be violated and spherical phase unwrapping might hence generate discontinuities between neighbouring points. Note that this investigation is only valid for all available positions $\boldsymbol{\Theta}_{\mathrm{rec}}$. The subset $\boldsymbol{\Theta}_{\mathrm{int}}$ is assumed to have a lower critical frequency due to larger angular distances between its neighbouring points.

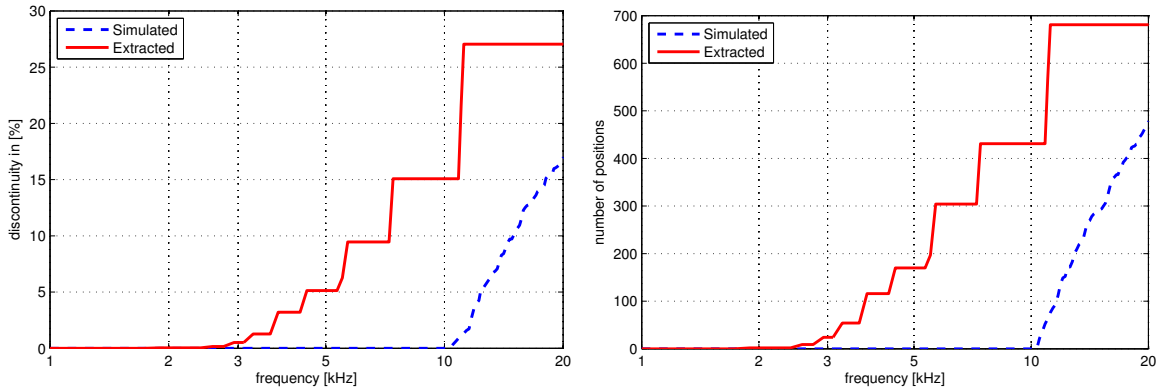


Figure 4.3.8: *Extracted and simulated linear phase.* Left: Mean discontinuity measure in [%] versus frequency. Right: Number of positions with any discontinuities versus frequency

Chapter 5

Spectral vs. Spherical Phase Unwrapping

5.1 Spatial Continuity

Mean discontinuity measure. In order to evaluate the spectral and spherical unwrapping procedures, resulting in the spectrally unwrapped phase Φ_{spectral} and the spherically unwrapped phase $\Phi_{\text{spherical}}$, the discontinuity measure E_{dis} was calculated for both unwrapped phase distributions (see *equation 4.3.1*). The discontinuity measure seems to be significantly lower when using the spherical unwrapping procedure (see *figure 5.1.1*). Since spherical unwrapping errors propagate along the spatial path, the discontinuity measure can be quite different for adjacent frequency bins. The spectral unwrapping procedure leads to a continuous increase of the discontinuity measure with frequency due to error propagation over frequency. The significant discontinuity measure value of the spectrally unwrapped phase for $f = 200\text{Hz}$ is caused by the discontinuity of the spectrally unwrapped phase distribution for $f = 170\text{Hz}$, see chapter 4.1).

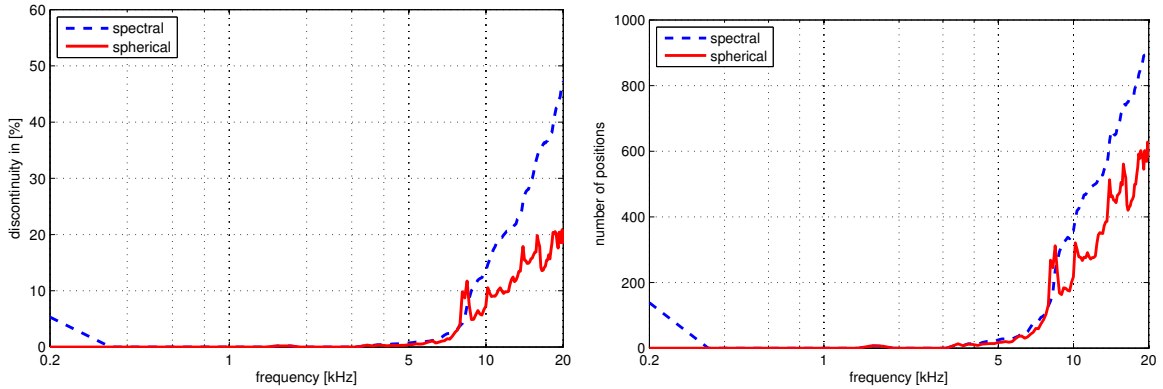


Figure 5.1.1: *Spectrally and spherically unwrapped phase*. Left: Mean discontinuity measure in [%] versus frequency. Right: Number of positions with any discontinuities versus frequency

Spatial distribution of $\boldsymbol{\Phi}_{\mathbf{spectral}}$, $\boldsymbol{\Phi}_{\mathbf{spherical}}$ and $\boldsymbol{E}_{\mathbf{dis}}$. Comparing spectrally and spherically unwrapped phase distributions and the resulting discontinuity measures for selected frequencies, the reasons for the apparent advantages of the spherical phase unwrapping algorithm become clear. While the spatial distribution of the spectrally unwrapped phase Φ_{spectral} is discontinuous for $f = 170\mathrm{Hz}$, the spherical phase unwrapping algorithm detects and eliminates this problem (see *figure 5.1.2*). In *figure 5.1.3*, the distributions of Φ_{spectral} and $\Phi_{\mathrm{spherical}}$ for $f = 5\mathrm{kHz}$ are depicted along with the corresponding discontinuity measures. The unwrapping methods produce a quite similar result for that frequency. With increasing frequency, depicted in *figures 5.1.4, 5.1.5, 5.1.6 and 5.1.7*, the differences between Φ_{spectral} and $\Phi_{\mathrm{spherical}}$ increase. The distributions of Φ_{spectral} show large regions of discontinuity, starting at ipsilateral positions. The discontinuities tend to propagate over frequency once they appear. The spherical phase unwrapping algorithm 1) forces the unwrapped phase distribution to a largely rotationally symmetric pattern around the interaural axis, 2) tends to produce most discontinuities at contralateral positions and 3) avoids error propagation over frequency.

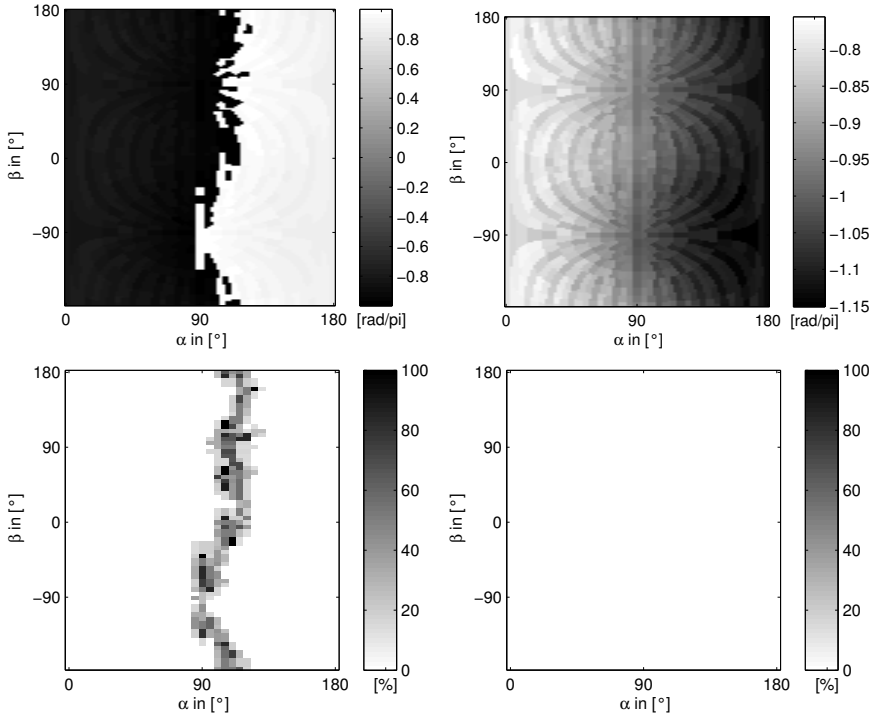


Figure 5.1.2: $f = 170\mathrm{Hz}$. top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

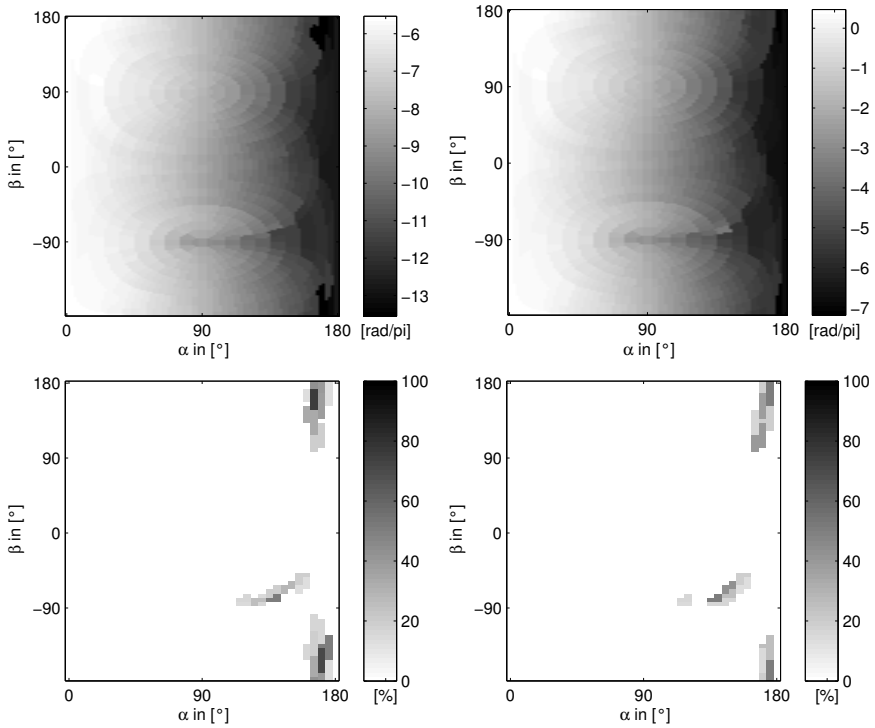


Figure 5.1.3: $f = 5\mathrm{kHz}$**.** top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

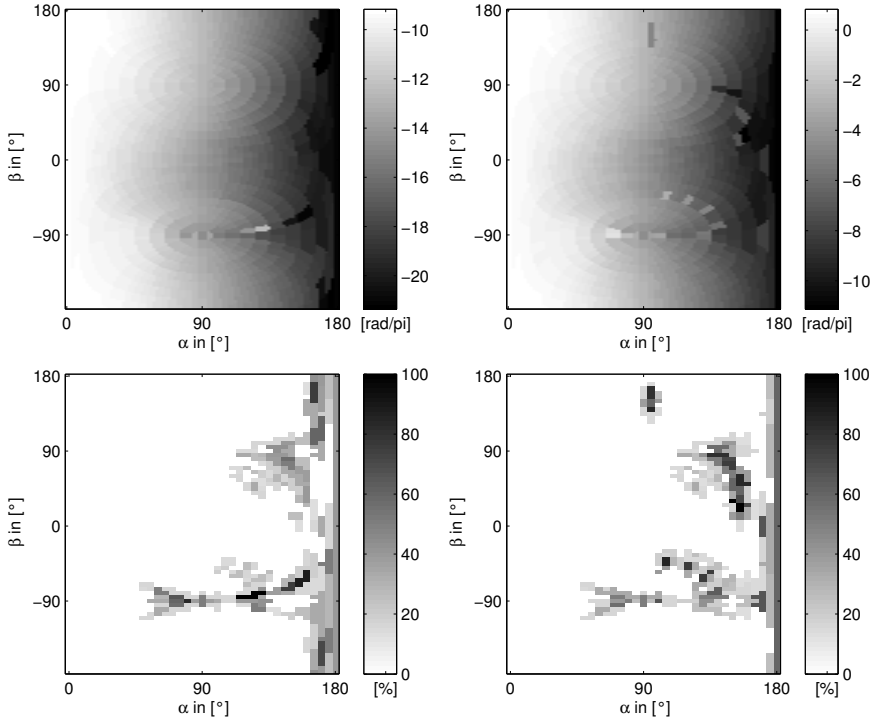


Figure 5.1.4: $f = 8\mathrm{kHz}$**.** top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

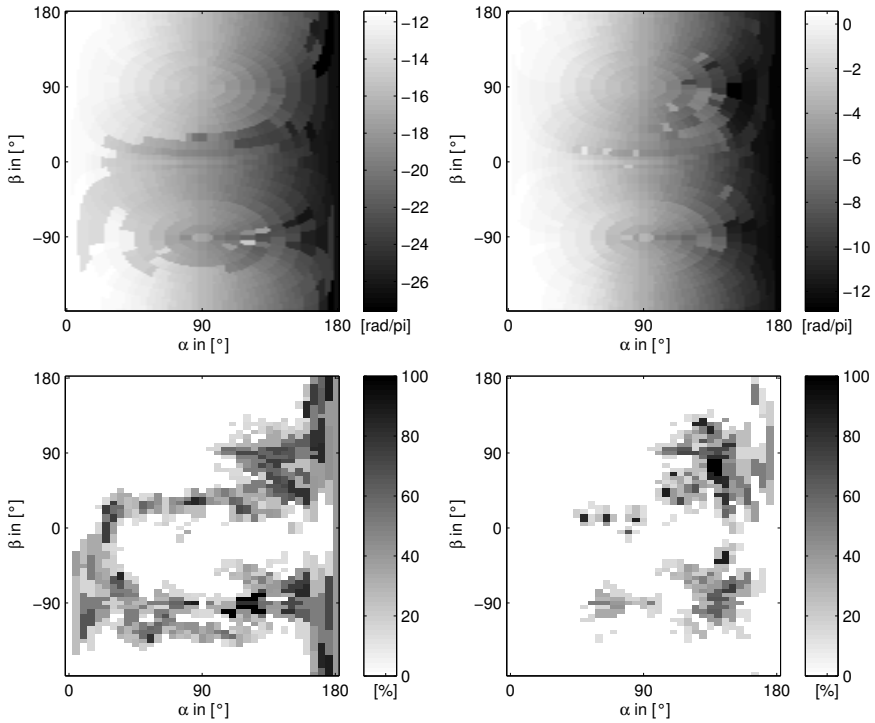


Figure 5.1.5: $f = 10\text{kHz}$**.** top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

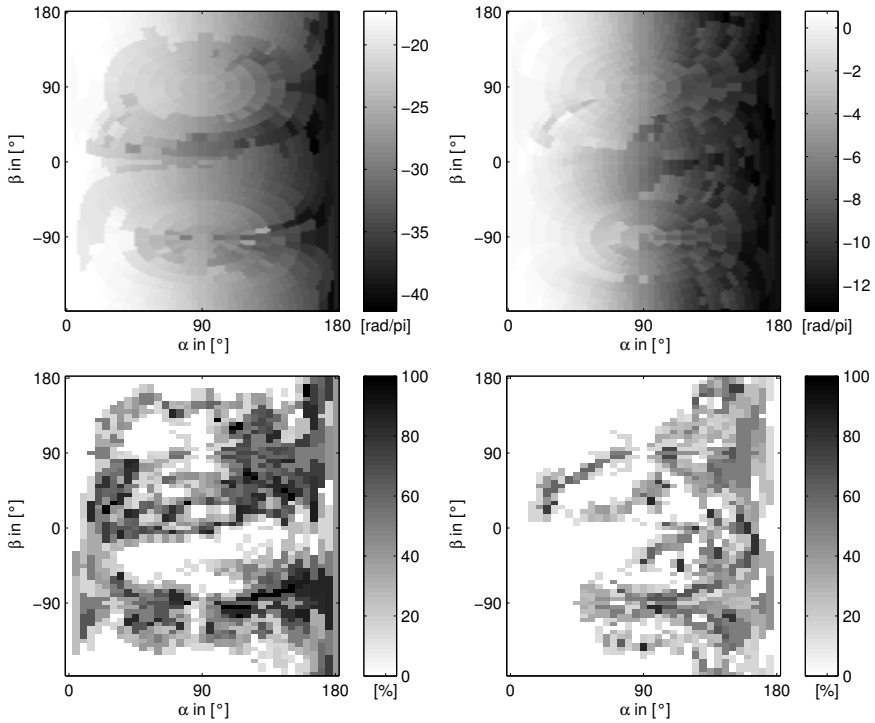


Figure 5.1.6: $f = 15\text{kHz}$**.** top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

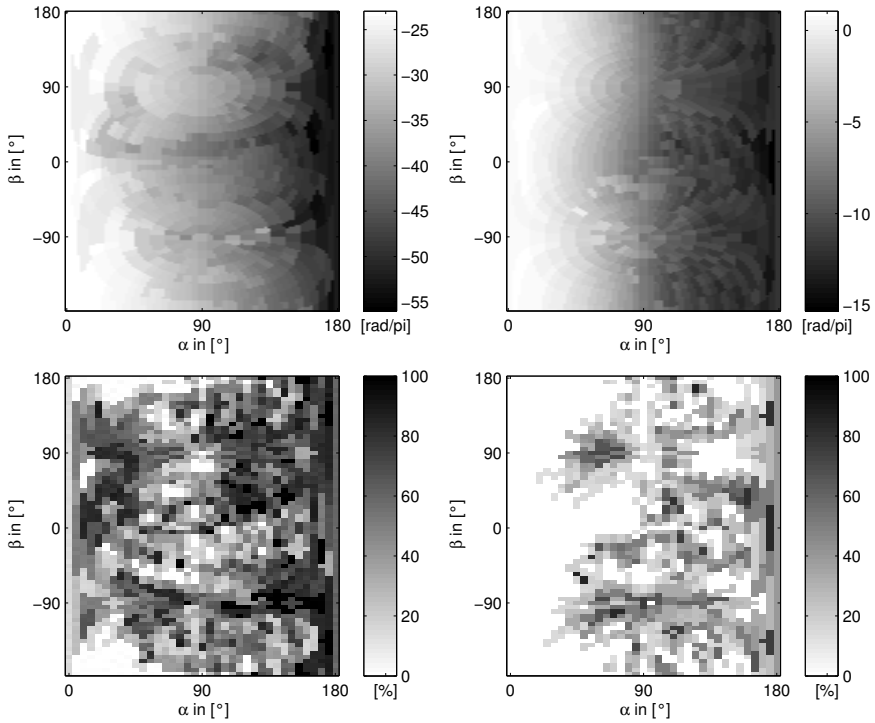


Figure 5.1.7: $f = 20\text{kHz}$. top: unwrapped phase, bottom: discontinuity measure E_{dis}. left: spectral phase unwrapping, right: spherical phase unwrapping.

Spatial range of unwrapped phase values and physical relevance. The range of the unwrapped phase distributions for frequencies larger than 8kHz is very different: The range of the spectrally unwrapped phase over all positions is approximately 15π for $f = 10\text{kHz}$, 23π for $f = 15\text{kHz}$ and 30π for $f = 30\text{kHz}$, all of which correspond to a maximal distance of $26cm$. This is a quite realistic value, considering a head radius of $r_{head} = 8.5cm$ and the corresponding distance from one ear to the other on the sphere $\mathrm{r_{head}}\ \pi \approx 27cm$. The range of the spherically unwrapped phase over all positions is much smaller, 13π for $f = 10\text{kHz}$, 13π for $f = 15\text{kHz}$ and 16π for $f = 30\text{kHz}$ which correspond to maximal distances of $22cm$, $15cm$ and $14cm$, respectively. The spherically unwrapped phase might hence not be physically meaningful for very high frequencies, where the phase is severely undersampled. Since this is basically a matter of 'inventing' the unwrapped phase in a spatial sense, it might just as well be adapted to some constraint, which is rotational symmetry and discontinuity reduction for ipsilateral positions in this case. Additionally, the phase of frequencies above 8kHz is quite irrelevant in terms of human perception.

Variance along the cone of confusion. In *figure 5.1.8* the maximal variances along the cone of confusion produced by both unwrapping procedures are depicted. Spherical phase unwrapping effectively minimizes those variances globally, while the local maxima of the discontinuity measure depicted in *figure 5.1.1* are clearly reflected as regions of relatively high variance. Furthermore, the low-frequency variance from the spectrally unwrapped phase in the

region of $90° < \alpha < 130°$ is eliminated by spherical unwrapping.

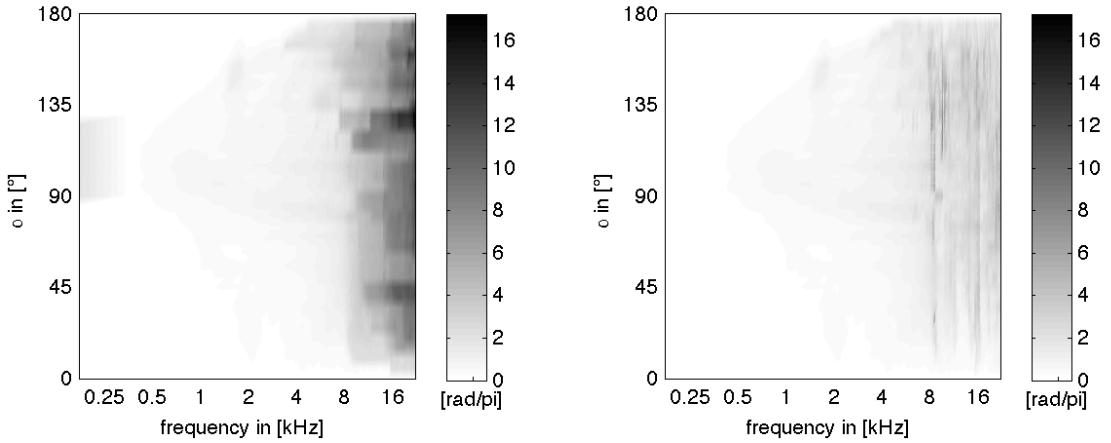


Figure 5.1.8: Maximal variance along the cone of confusion. Left: Spectrally unwrapped phase, right: Spherically unwrapped phase

5.2 Spherical Harmonic Transform Performance

In this section, the suitability of the spectral and spherical phase unwrapping procedures for the DSHT is compared using the phase error measure E_ϕ.

Phase error definition. The phase error is a measure of the similarity between original phase ϕ_{ref} and manipulated phase ϕ. It is defined as

$$E_\phi(\boldsymbol{\theta}) = \frac{100}{\pi} \frac{\sum_{l=0}^{\mathrm{L}-1} \mathcal{W}(|\phi_{\mathrm{ref}}[\boldsymbol{\theta}, l] - \Phi[\boldsymbol{\theta}, l]|)\ \nu[l]}{\sum_{l=0}^{\mathrm{L}-1} \nu[l]}\ [\%], \tag{5.2.1}$$

with the frequency dependent weights

$$\nu[l] = \begin{cases} \frac{1}{f[l]}, & \text{if } f[l] > 1\mathrm{kHz} \\ 1, & \text{if } f[l] \leq 1\mathrm{kHz}, \end{cases}$$

L denotes the number of the frequency bins l and $\mathcal{W}$ is the phase wrapping operator from *equation 2.3.2*.

DSHT. The DSHT was calculated for the reconstruction grid $\boldsymbol{\Theta}_{\mathrm{rec}}$ and for the interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$, depicted in *figure 3.6.1*. The DSHT was applied as described in chapter 3.8, afterwards the phase error E_ϕ was calculated.

Mean phase error. The mean phase error is depicted in *figure 5.2.1*. Obviously, the spherically unwrapped phase is more suitable for the reconstruction than the spectrally unwrapped phase. The mean phase errors produced by interpolation are similar for both unwrapping methods. This is possibly due to a degraded performance of the spherical phase unwrapping algorithm using the reduced interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$ because of a violation of the Itoh-Condition for comparatively low frequencies.

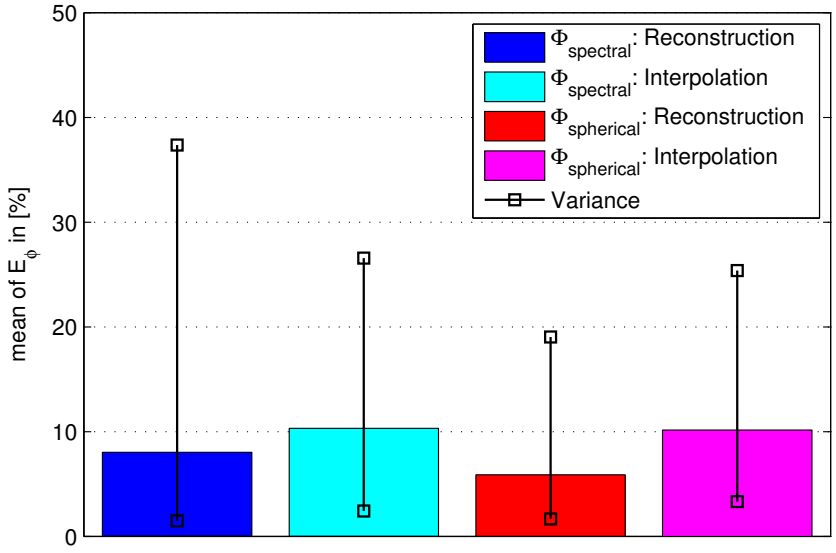
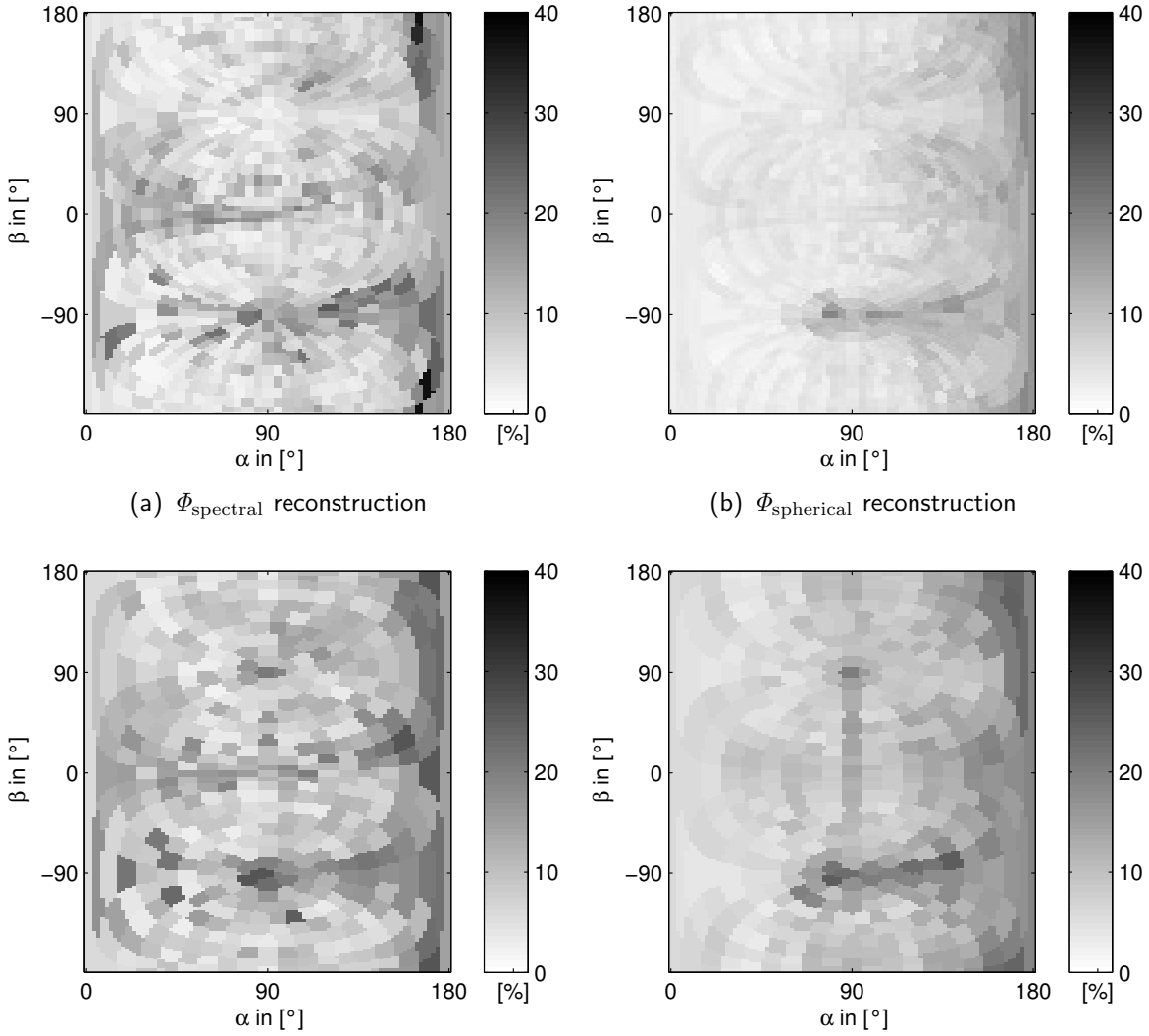


Figure 5.2.1: Mean phase error

Phase error distribution. The distributions of the phase errors are depicted in *figure 5.2.2*. In accordance with the distribution of spatial phase discontinuities (see chapter 5.1), the phase errors resulting from reconstruction and interpolation of the spherically unwrapped phase are relatively low for ipsilateral positions. Maximal errors occur in the region of the torso and in extreme contralateral positions. The phase errors resulting from reconstruction and interpolation of the spectrally unwrapped phase show the same tendency, while additional significant errors in ipsilateral positions can be observed. This also corresponds to the more or less random distribution of the spatial phase discontinuities produced by spectral phase unwrapping.

Conclusions. The spherical phase unwrapping algorithm seems to provide a higher suitability for the DSHT in general. Nevertheless, the results suggest that it is less robust to undersampling of the phase distribution, i.e. violations of the Itoh-Condition. This is not surprising, since spherical unwrapping progresses from one spatial position to another, while spectral unwrapping progresses from one frequency bin to another. On the one hand, spectral phase unwrapping does not account for spatial consistency, which can lead to unnecessary spatial phase discontinuities in the appropriately sampled case. On the other hand, if the

Itoh-Condition is violated, which is clearly the case for higher frequencies, spectral unwrapping might be more robust. Still, the spatial consistency of spectral phase unwrapping is random and can easily be corrupted by errors due to the measurement itself and post-measurement processing.



(a) Φ_{spectral} reconstruction

(b) $\Phi_{\mathrm{spherical}}$ reconstruction

(c) Φ_{spectral} interpolation

(d) $\Phi_{\mathrm{spherical}}$ interpolation

Figure 5.2.2: Phase error distribution

Chapter 6

Results

In this chapter, the performance of the DSHT using different kinds of HRTF representations is evaluated. The evaluation is performed for reconstruction via the DSHT using the reconstruction grid $\boldsymbol{\Theta}_{\mathrm{rec}}$ and for interpolation via the DSHT using the interpolation grid $\boldsymbol{\Theta}_{\mathrm{int}}$. The sampling grids are depicted in *figure 3.6.1*. The definitions of the maximal orders $\mathrm{N}_{\mathrm{rec}}$ and $\mathrm{N}_{\mathrm{int}}$, corresponding to the respective sampling grids, have been derived in chapter 3.6 according to the invertibility of the resulting spherical harmonic matrices $\boldsymbol{Y}_{\mathrm{N}_{\mathrm{rec}}}$ and $\boldsymbol{Y}_{\mathrm{N}_{\mathrm{int}}}$.

The HRTF representations used for comparison are

I. HRIRs $h[\boldsymbol{\theta}, k]$,
II. HRTF magnitude $|H[\boldsymbol{\theta}, \omega_l]|$ and *spectrally* unwrapped phase $\Phi_{\mathrm{spectral}}[\boldsymbol{\theta}, \omega_l]$,
III. HRTF magnitude $|H[\boldsymbol{\theta}, \omega_l]|$ and *spherically* unwrapped phase $\Phi_{\mathrm{spherical}}[\boldsymbol{\theta}, \omega_l]$.

The evaluation of HRTFs modified by spatial manipulation via the DSHT is psychoacoustically motivated, the respective error measures are defined in the corresponding sections.

6.1 ITD Error

ITD error definition. The ITD is the localization cue that determines the lateral position of a sound source perceived by listeners for frequencies below $1.3\mathrm{kHz}$ [MM02]. The ITD error is defined as

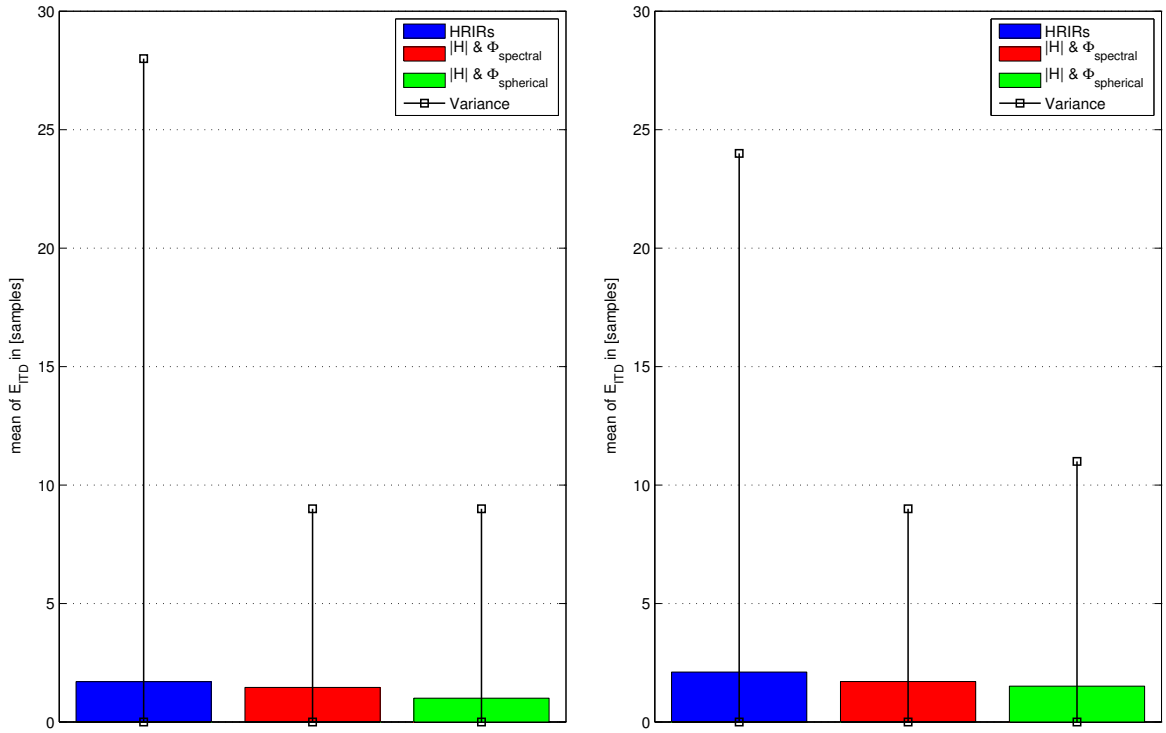
$$E_{\mathrm{ITD}}(\boldsymbol{\theta}) = |ITD_{\mathrm{ref}}(\boldsymbol{\theta}) - ITD(\boldsymbol{\theta})| \text{ in } [\mathrm{samples}], \tag{6.1.1}$$

where

$$\begin{aligned} ITD_{\mathrm{ref}}(\boldsymbol{\Theta}) &= \tau_{\mathrm{ref,R}}(\boldsymbol{\theta}) - \tau_{\mathrm{ref,L}}(\boldsymbol{\theta}), \\ ITD(\boldsymbol{\theta}) &= \tau_{\mathrm{R}}(\boldsymbol{\theta}) - \tau_{\mathrm{L}}(\boldsymbol{\theta}). \end{aligned}$$

The onset delays $\tau_{\mathrm{ref,R}}$ and $\tau_{\mathrm{ref,L}}$ are extracted from the original left and right impulse responses $h_{\mathrm{ref,R}}$ and $h_{\mathrm{ref,L}}$ and τ_{R} and τ_{L} from the manipulated left and right impulse responses h_{R} and h_{L} (see annex, chapter 8.2 for onset delay extraction).

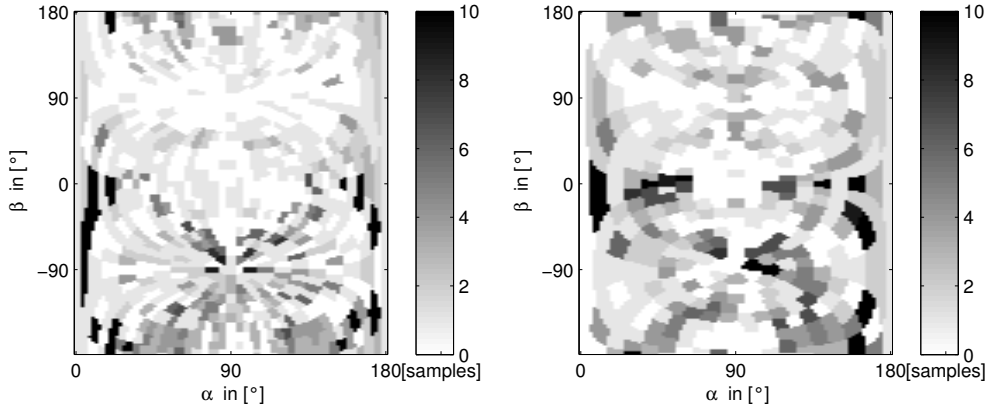
Mean ITD error. The mean of the ITD errors produced by reconstruction of the three different HRTF representations shows a clear tendency. The reconstruction of HRIRs seems to perform worst, with a mean error of $1.7[\mathrm{sampes}] \approx 38\mu\mathrm{s}$. This is due to the 'pre-delays' resulting from the DSHT of temporally displaced impulse responses, which have already been observed in chapter 3.8.1 and also account for the high maxima of almost $30[\mathrm{samples}]$. Reconstruction of $|H|$ and Φ_{spectral} results in a mean error of $1.5[\mathrm{samples}] \approx 34\mu\mathrm{s}$, while reconstruction of $|H|$ and $\Phi_{\mathrm{spherical}}$ produces the lowest mean error of $1[\mathrm{sample}] \approx 23\mu\mathrm{s}$. The difference between reconstruction of spectral and spherical unwrapping can be explained by the low-frequency phase discontinuity produced by spectral phase unwrapping, which has been eliminated using the spherical phase unwrapping algorithm (see *figure 5.1.2*).
Unsurprisingly, the mean ITD errors produced by interpolation are generally higher, while the ranking remains the same. The difference between the spherical and spectral phase unwrapping methods is smaller compared to reconstruction. This is due to the spherical phase unwrapping algorithm's lack of robustness against spatial undersampling observed in chapter 5.2.



(a) Reconstruction (b) Interpolation

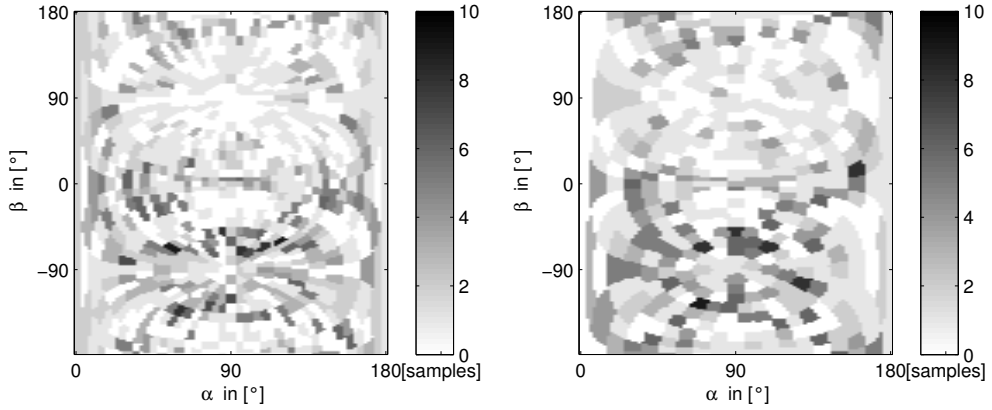
Figure 6.1.1: Mean ITD errors

ITD error distribution. The most significant ITD errors produced by interpolation and reconstruction of all three HRTF representation mainly occur in the lower hemisphere. Using HRIRs, extreme lateral positions are particularly affected due to significant 'pre-delays'. The spherically unwrapped phase clearly yields lower errors compared to the spectrally unwrapped phase, especially in the upper hemisphere.



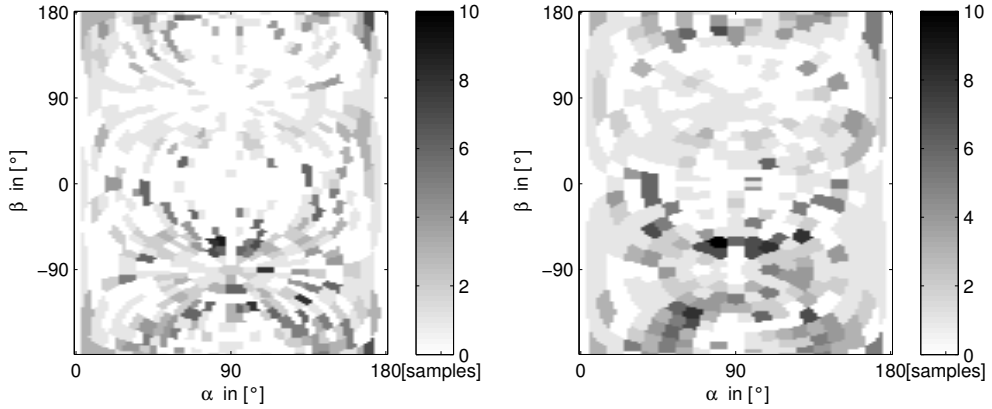
(a) HRIR reconstruction

(b) HRIR interpolation



(c) $|H|$ & Φ_{spectral} reconstruction

(d) $|H|$ & Φ_{spectral} interpolation



(e) $|H|$ & $\Phi_{\mathrm{spherical}}$ reconstruction

(f) $|H|$ & $\Phi_{\mathrm{spherical}}$ interpolation

Figure 6.1.2: ITD error distributions

6.2 ILD Error

ILD error definition. The ILD (interaural level difference) is the localization cue that determines the lateral position of a sound source perceived by listeners for frequencies above $1.3\mathrm{kHz}$ [MM02]. The ILD can be defined as

$$ILD(\boldsymbol{\theta}, b) = 20\log(|H_{\mathrm{R}}[\boldsymbol{\theta}, b]|) - 20\log(|H_{\mathrm{L}}[\boldsymbol{\theta}, b]|),$$

where the bark bands b are defined as

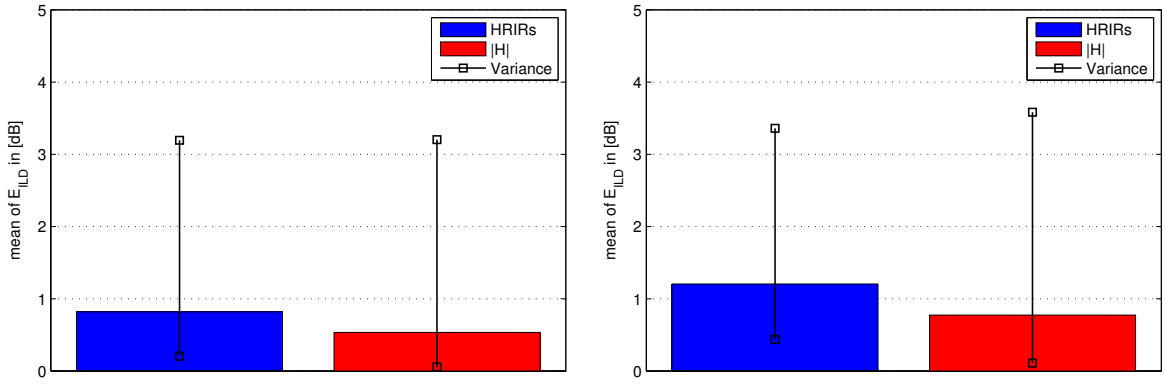
$$b = 13\arctan(0.00076f) + 3.5\arctan\left(\frac{f^2}{7500^2}\right).$$

The ILD error is calculated as the mean of the absolute difference between original ILD_{ref} and manipulated ILD

$$E_{\mathrm{ILD}}(\boldsymbol{\theta}) = \frac{1}{\mathrm{B}}\sum_{b=1}^{\mathrm{B}} |ILD_{\mathrm{ref}}(\boldsymbol{\theta}, b) - ILD(\boldsymbol{\theta}, b)| \text{ in } [\mathrm{dB}], \tag{6.2.1}$$

where B denotes the number of the discrete bark bands b.

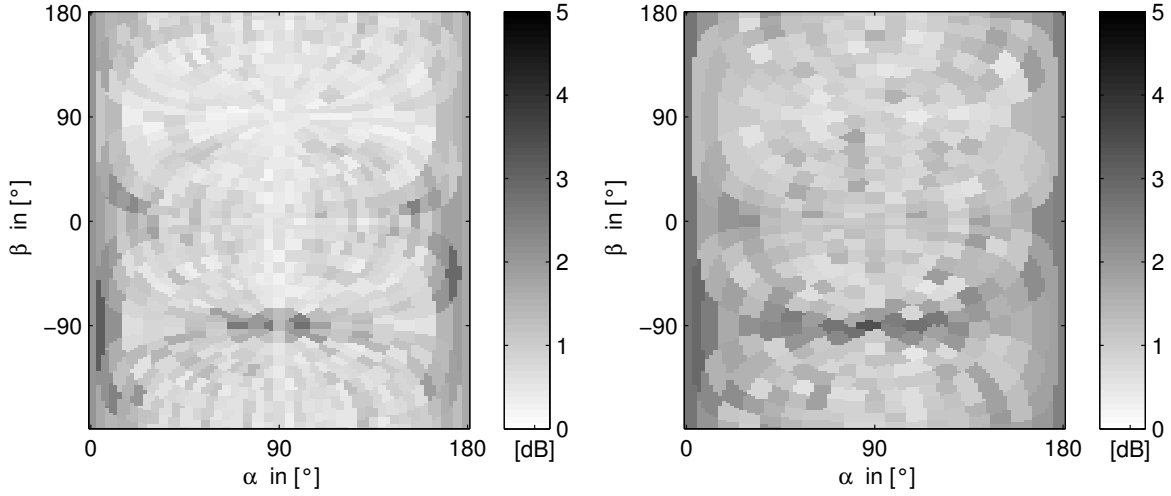
Mean ILD error. *Figure 6.2.1* indicates that the direct reconstruction and interpolation of the magnitude responses $|H[\boldsymbol{\theta}, \omega_l]|$ is significantly superior to indirect reconstruction and interpolation using the HRIRs in terms of the ILD error. The errors produced by interpolation are higher than the errors produced by reconstruction, which is due to the physical complexity of the true magnitude distribution.



(a) Reconstruction (b) Interpolation

Figure 6.2.1: Mean ILD errors

ILD error distribution. The distributions of the ILD error show a similar tendency: Maximal errors generally occur at extreme lateral positions and below the head, in the region of the torso. The mean errors depicted in *figure 6.2.1* are reflected as the overall error level in the distributions. The errors for interpolation are higher than for reconstruction. The direct reconstruction and interpolation of the magnitude leads to lower ILD errors than its time-domain counterpart, the reconstruction and interpolation of HRIRs.



(a) HRIR reconstruction

(b) HRIR interpolation

(c) $|H|$ reconstruction

(d) $|H|$ interpolation

Figure 6.2.2: ILD error distributions

6.3 Magnitude Error

Magnitude error definition. The magnitude error used in this thesis was introduced by Minnaar et al. based on the results of a psychoacoustical experiment [MPC05]. It is defined as

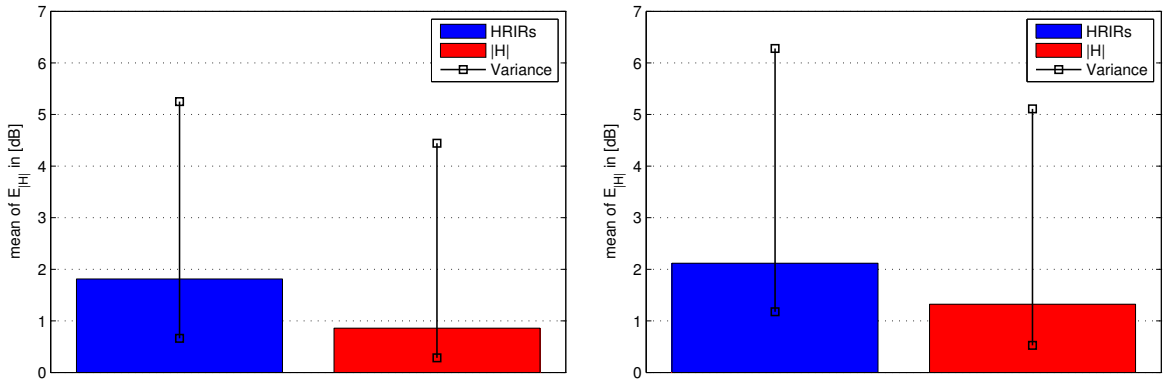
$$E_{|\mathrm{H}|}(\boldsymbol{\theta}) = \frac{1}{\mathrm{p}} \sum_{l=l_{\log,1}}^{l_{\log,\mathrm{p}}} |E_{|\mathrm{H}|,\mathrm{R}}[\boldsymbol{\theta}, l] + E_{|\mathrm{H}|,\mathrm{L}}[\boldsymbol{\theta}, l]| \text{ in } [\mathrm{dB}], \tag{6.3.1}$$

where $l_{\log,1}$, ..., $l_{\log,\mathrm{p}}$ are p frequency bins, logarithmically distributed between $100\mathrm{Hz}$ and $20\mathrm{kHz}$, and

$$\begin{aligned} E_{|\mathrm{H}|,\mathrm{L}}[\boldsymbol{\theta}, l] &= |20\log(|H_{\mathrm{ref,L}}[\boldsymbol{\theta}, l]|) - 20\log(|H_{\mathrm{L}}[\boldsymbol{\theta}, l]|)|, \\ E_{|\mathrm{H}|,\mathrm{R}}[\boldsymbol{\theta}, l] &= |20\log(|H_{\mathrm{ref,R}}[\boldsymbol{\theta}, l]|) - 20\log(|H_{\mathrm{R}}[\boldsymbol{\theta}, l]|)|. \end{aligned}$$

$|H_{\mathrm{ref,R}}|$ and $|H_{\mathrm{ref,R}}|$ denote the original left and right magnitude spectra, $|H_{\mathrm{R}}|$ and $|H_{\mathrm{L}}|$ the manipulated left and right magnitude spectra.
Minnaar et al. found that the just-noticeable difference (JND) in terms of this error measure is $1\mathrm{dB}$. In other words, if $E_{|\mathrm{H}|}(\boldsymbol{\theta}) \leq 1\mathrm{dB}$, listeners could not detect the difference between $|H_{\mathrm{R}}|$ and $|H_{\mathrm{L}}|$ and the original magnitude spectra $|H_{\mathrm{ref,R}}|$ and $|H_{\mathrm{ref,L}}|$.

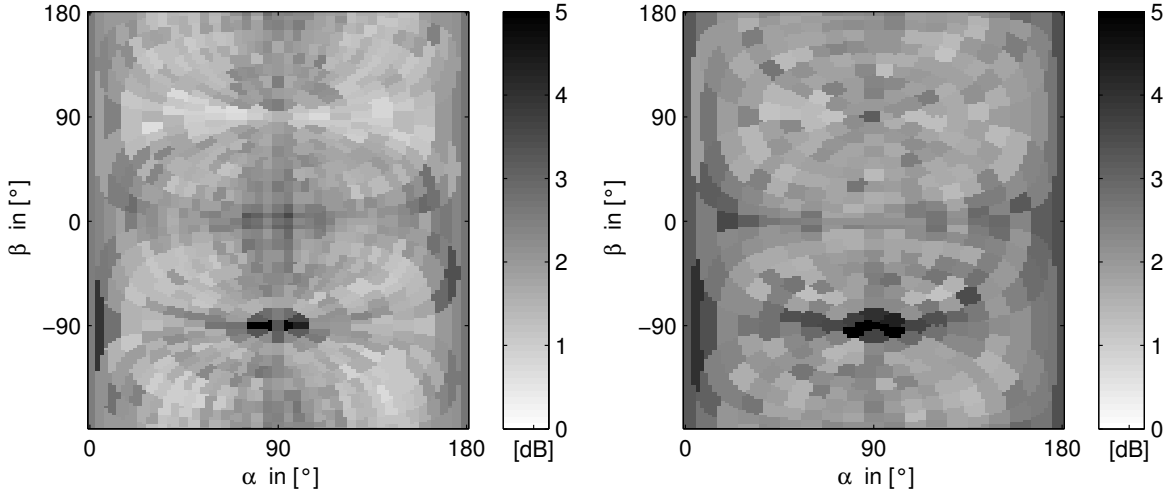
Mean magnitude error. The values of the magnitude error depicted in *figure 6.3.2* naturally correspond to the ILD error, which also depends on the spectral magnitude. In the reconstruction case, the direct DSHT of the magnitude produces a mean error below the JND of $1\mathrm{dB}$, while the DSHT of HRIRs leads to a mean error of almost $2\mathrm{dB}$. The mean errors produced by interpolation of both HRTF representations increase compared to reconstruction, while the ratio between DSHT of magnitude and DSHT of HRIRs is similar.



(a) Reconstruction (b) Interpolation

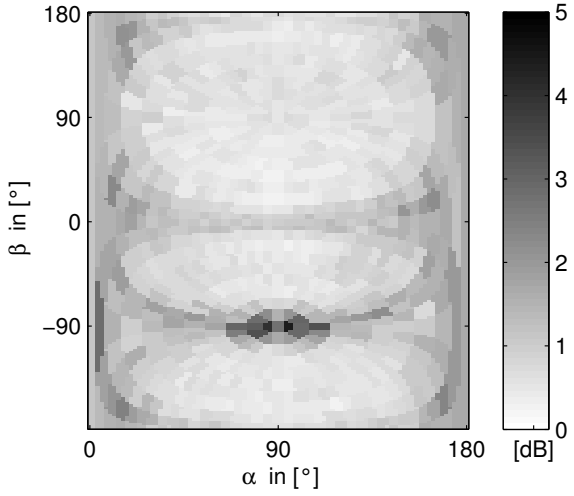
Figure 6.3.1: Mean magnitude errors

Magnitude error distribution. The distributions of the magnitude errors largely correspond to the distributions of the ILD errors, which are depicted in *figure 6.2.2*. Maximal errors generally occur at extreme lateral positions, in the region of the torso and in the horizontal plane. The reconstruction of HRIRs produces additional high errors around the median plane. Since ILD and ITD do not provide meaningful localization information in the median plane, this can lead to significant problems concerning the detection of sound source positions.

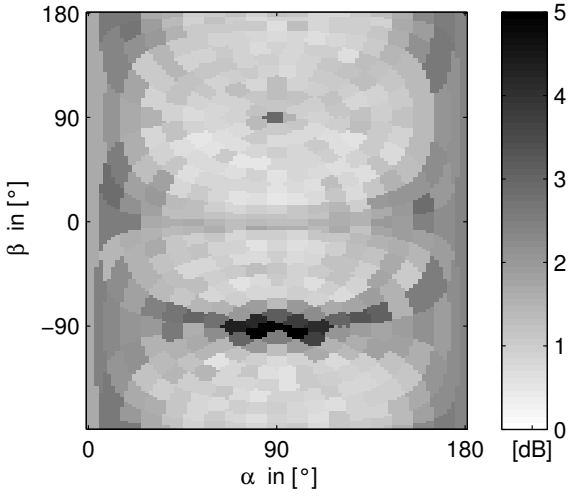


(a) HRIR reconstruction

(b) HRIR interpolation



(c) $|H|$ reconstruction



(d) $|H|$ interpolation

Figure 6.3.2: Magnitude error distributions

6.4 Most Important Basis Functions

In this section, the importance of the respective spherical harmonics for HRIRs, magnitude $|H|$, spectrally unwrapped phase Φ_{spectral}, and spherically unwrapped phase $\Phi_{\mathrm{spherical}}$ is investigated. Assuming a similar behaviour for left and right HRTFs due to the symmetric shape of the dummy head used for the HRTF measurements, only the left ear HRTFs are investigated here.

Nomenclature. The spherical harmonic coefficients were calculated for the HRIRs $h[\boldsymbol{\theta}, k]$, the magnitude responses $|H[\boldsymbol{\theta}, \omega_l]|$, the spectrally unwrapped phase responses $\Phi_{\mathrm{spectral}}[\boldsymbol{\theta}, \omega_l]$ and the spherically unwrapped phase responses $\Phi_{\mathrm{spherical}}[\boldsymbol{\theta}, \omega_l]$ using all positions $\boldsymbol{\Theta}_{\mathrm{rec}}$

$$\begin{aligned}
\boldsymbol{\gamma}_{\mathrm{N_{rec}},\mathrm{h}}[k] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{rec}}}\ h[\boldsymbol{\Theta}_{\mathrm{rec}}, k], \\
\boldsymbol{\gamma}_{\mathrm{N_{rec}},|\mathrm{H}|}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{rec}}}\ |H[\boldsymbol{\Theta}_{\mathrm{rec}}, \omega_l]|, \\
\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi_{\mathrm{spectral}}}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{rec}}}\ \Phi_{\mathrm{spectral}}[\boldsymbol{\Theta}_{\mathrm{rec}}, \omega_l], \\
\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi_{\mathrm{spherical}}}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{rec}}}\ \Phi_{\mathrm{spherical}}[\boldsymbol{\Theta}_{\mathrm{rec}}, \omega_l],
\end{aligned}$$

and using the subset of positions $\boldsymbol{\Theta}_{\mathrm{int}}$

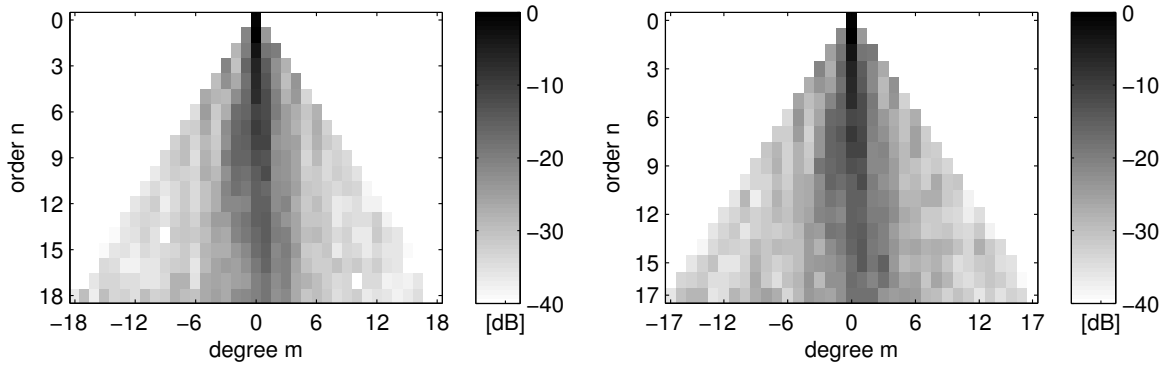
$$\begin{aligned}
\boldsymbol{\gamma}_{\mathrm{N_{int}},\mathrm{h}}[k] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{int}}}\ h[\boldsymbol{\Theta}_{\mathrm{int}}, k], \\
\boldsymbol{\gamma}_{\mathrm{N_{int}},|\mathrm{H}|}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{int}}}\ |H[\boldsymbol{\Theta}_{\mathrm{int}}, \omega_l]|, \\
\boldsymbol{\gamma}_{\mathrm{N_{int}},\Phi_{\mathrm{spectral}}}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{int}}}\ \Phi_{\mathrm{spectral}}[\boldsymbol{\Theta}_{\mathrm{int}}, \omega_l], \\
\boldsymbol{\gamma}_{\mathrm{N_{int}},\Phi_{\mathrm{spherical}}}[\omega_l] &= \boldsymbol{Y}^{\dagger}_{\mathrm{N_{int}}}\ \Phi_{\mathrm{spherical}}[\boldsymbol{\Theta}_{\mathrm{int}}, \omega_l].
\end{aligned}$$

Calculation of maximal coefficients. Since the DSHT is applied to each frequency bin or each sample separately, the calculation of the overall importance of a specific coefficient over all frequencies or all samples requires some consideration. For the coefficients derived from the magnitude, the spectrally and spherically unwrapped phase, the absolute values were normalized according to their maximum value per frequency bin and weighted with $1/f$ for frequencies greater than $500\mathrm{Hz}$ to avoid an overestimation of high-frequency components. For the coefficients derived from the HRIRs, the coefficients were filtered according to the $1/f$-weighting; afterwards, the absolute values were normalized according to their overall maximum per sample and position, thus emphasizing the importance of high-amplitude components. The maxima of the weighted coefficients over time or frequency are depicted in the following.

Rotated spherical harmonics. Note that the DSHT is applied using the polar plane angle β and the lateral angle α. The basis functions are hence rotated by $90°$ in comparison to the commonly used DSHT with azimuth and elevation angles. For a more detailed description of

this modification of the DSHT, the interested reader is referred to chapter 3.

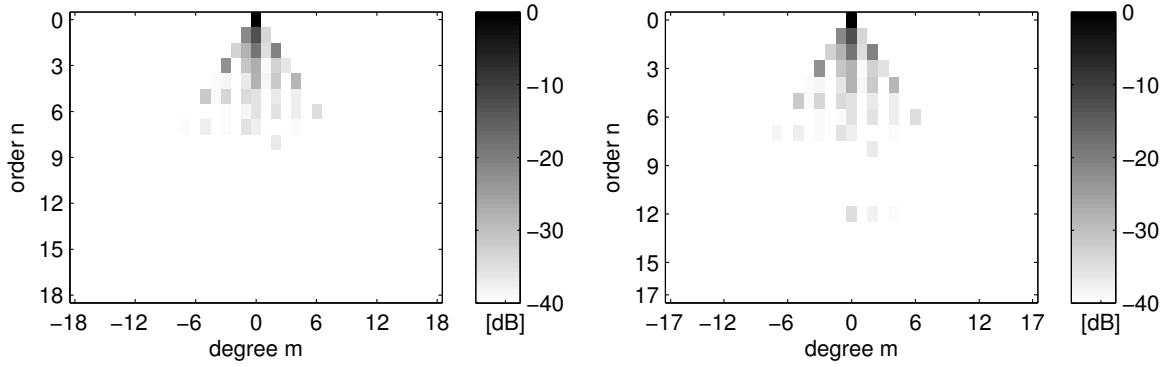
DSHT of HRIRs. The maximal spherical harmonic coefficients derived from the HRIRs indicate a rather high importance of almost all available basis functions. This reflects the high spatial complexity of the impulse responses. Due to the rotation of the basis functions, the highest coefficient values occur for $m = 0$. This is due to the almost rotationally symmetric distribution of the impulse responses' energy in relation to the interaural axis.



(a) $\max(\boldsymbol{\gamma}_{\mathrm{N_{rec,h}}}[k])$, reconstruction grid (b) $\max(\boldsymbol{\gamma}_{\mathrm{N_{int,h}}}[k])$, interpolation grid

Figure 6.4.1: Maximal weighted coefficients derived from DSHT of HRIRs

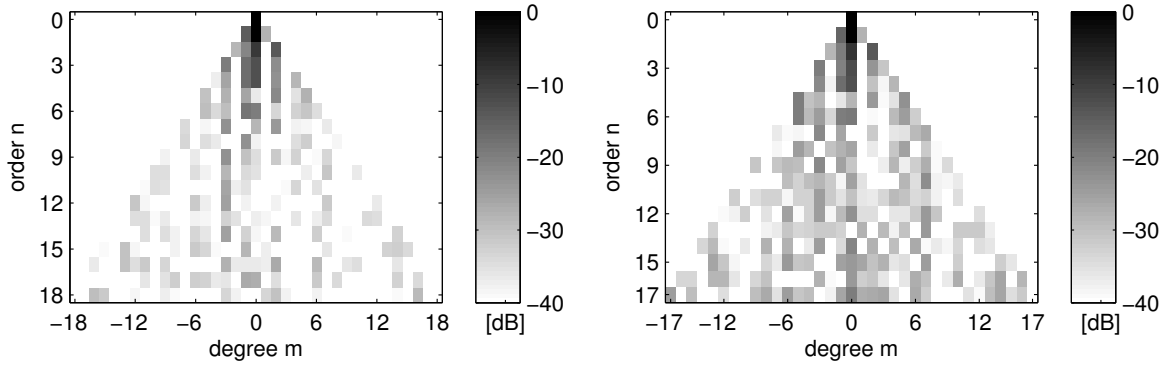
DSHT of magnitude responses. The maximal spherical harmonic coefficients derived from the HRTFs' magnitude $|H|$ show very low significance for higher-order basis functions. Consequently, the maximal order could possibly be set to $\mathrm{N} = 5$ without a significant loss of reconstruction and interpolation accuracy.



(a) $\max(\boldsymbol{\gamma}_{\mathrm{N_{rec,|H|}}}[\omega_l])$, reconstruction grid (b) $\max(\boldsymbol{\gamma}_{\mathrm{N_{int,|H|}}}[\omega_l])$, interpolation grid

Figure 6.4.2: Maximal weighted coefficients derived from DSHT of magnitude responses

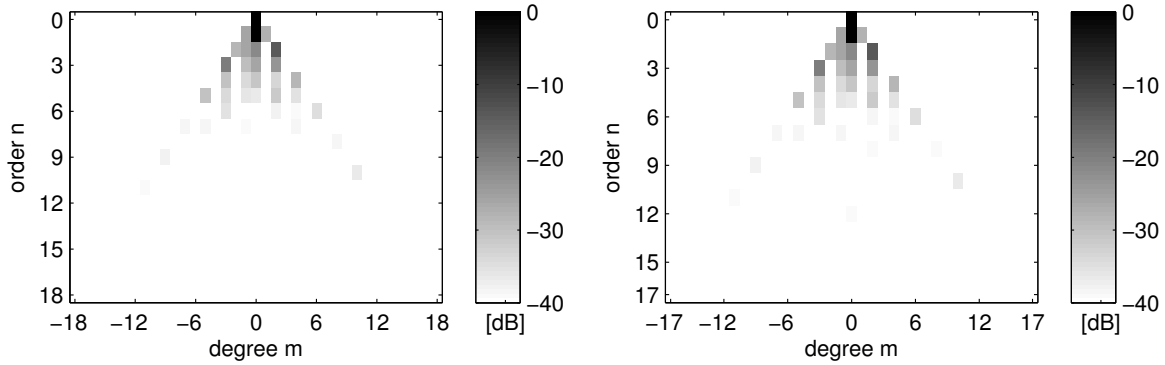
DSHT of *spectrally* unwrapped phase responses. The maximal coefficients derived from the HRTFs' spectrally unwrapped phase responses reflect a rather high spatial complexity of the unwrapped phase distribution. While the highest density of most important basis functions can be found for low orders n and degrees m, coefficients for higher orders and degrees are not negligible, especially when using the coarser interpolation sampling grid.



(a) $\max(\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi_{\mathrm{spectral}}}[\omega_l])$, reconstruction grid (b) $\max(\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi_{\mathrm{spectral}}}[\omega_l])$, interpolation grid

Figure 6.4.3: Maximal weighted coefficients derived from DSHT of spectrally unwrapped phase responses

DSHT of *spherically* unwrapped phase responses. As opposed to the most important basis functions derived from the spectrally unwrapped phase, the maximal coefficients depicted in *figure 6.4.4* show very low significance for higher-order basis functions. This reflects a lower spatial complexity of the spherically unwrapped phase distribution compared to the spectrally unwrapped phase distribution. Most significant coefficient values occur for Y_0^0, which represents the constant component, and for Y_1^0, which represents the simplest rotationally symmetric pattern with respect to the interaural axis.



(a) $\max(\boldsymbol{\gamma}_{\mathrm{N_{rec}},\Phi_{\mathrm{spherical}}}[\omega_l])$, reconstruction grid (b) $\max(\boldsymbol{\gamma}_{\mathrm{N_{int}},\Phi_{\mathrm{spherical}}}[\omega_l])$, interpolation grid

Figure 6.4.4: Maximal weighted coefficients derived from DSHT of spherically unwrapped phase responses

Since the basis of both unwrapping methods is the same wrapped phase ϕ, the spherical phase unwrapping algorithm can be assumed to effectively minimize the spatial complexity of the phase distribution. This confirms the results of the investigation of spherical and spectral phase unwrapping methods in chapter 5. The spherically unwrapped phase consequently allows for a significant reduction of the maximal order to $\mathrm{N} = 5$.

Chapter 7

Conclusion

7.1 Summary of the results

In this thesis, the suitability of different HRTF representations for the discrete spherical harmonic transform has been investigated, particularly concerning interpolation quality. Contrary to the commonly used minimum phase and delay approximation of HRTFs, the original unmodified HRTFs have been used. It has been found that a separate interpolation of HRTFs represented by their spectral magnitude and unwrapped phase responses is in principle superior to an interpolation of their temporally unaligned time domain representations (HRIRs). Since the phase responses are only available as discontinuous functions, wrapped to the interval $]-\pi, \pi]$, phase unwrapping had to be applied in order to reconstruct the absolute phase. Phase unwrapping is a very delicate task and has therefore been further investigated. It has been found that spectral phase unwrapping, which is typically used in audio applications, does not necessarily provide meaningful results concerning the required spatial phase continuity. An alternative method has been developed, which unwraps the phase in a spatial sense on the sphere, based on geometric relations between the discrete sampling positions. Results suggest that the spherical phase unwrapping algorithm presented in this thesis effectively minimizes phase discontinuities on the sphere. Additionally, the application of spherical phase unwrapping significantly simplifies the discrete spherical harmonic spectrum. An investigation of the spherical harmonics' relevance suggests that a set of HRTFs measured on a 1014-point spatial sampling grid could be reducible to 36 spherical harmonic coefficients without significant loss of accuracy using the proposed method.

7.2 Interpretation and Outlook

Application of spherical phase unwrapping for coarse spatial sampling grids. While the representation of HRTFs as their magnitude and spherically unwrapped phase responses allows for very accurate interpolation in principle, the quality of spherical phase unwrapping is heavily dependent on the relation between spatial resolution and frequency. The true absolute phase must not change more than π between neighbouring points in space to guarantee accurate phase unwrapping. Using typically applied spatial sampling grids for HRTF measurements, this condition is violated for smaller wavelengths, i.e. higher frequencies. As the measurement duration is usually very long, which furthermore leads to problems with the fixation of human subjects, a reduction of the measurement grid is desirable. Unfortunately, a reduced spatial resolution leads to a decrease of the critical frequency for accurate spherical phase unwrapping.

Two approaches to solving this conflict shall be mentioned here:

1. To determine the position of a sound source, the human auditory system resolves interaural time differences (i.e. phase differences) only for low frequencies. Above $1.3\mathrm{kHz}$ [MM02], the phase can therefore be replaced by an excess phase and minimum phase approximation without loss of localization accuracy. A perceptually motivated approach can hence be realized by using the spherical phase unwrapping algorithm only for low frequencies, where the spatial resolution can be assumed to be sufficient even if it is coarse. Phase values for higher frequencies can be replaced by a) the interpolated onset delay, yielding the excess phase, and b) the minimum phase response calculated from the interpolated magnitude response.

2. It has been observed that the spherical phase unwrapping algorithm fails to produce physically meaningful results if the phase is spatially undersampled. Phase-aliasing suppression is already implemented in the algorithm, forcing the phase to globally decrease from the ipsilateral to the contralateral side. Nevertheless, it might be beneficial to extend the algorithm using a physically motivated estimation of the absolute phase, e.g. via onset delay simulation by means of a geometric model. This approach might merge the advantages of spherical and spectral phase unwrapping, taking the spatial *and* the spectral consistency of the phase into account.

Considerations on spatial sampling strategies. During the work on this thesis, many different sets of HRTFs have been investigated. One of them has been simulated using a magnetic resonance imaging (MRI) scanned BEM modeling grid of a Neumann head, as documented in [GK07]. The boundary element simulation was conducted at LIMSI/CNRS, Orsay, using an implementation by courtesy of Makoto Otani, see [OISI10]. The sampling grid was defined using a 6° equidistant angular spacing of lateral angles and polar plane angles, depicted in *figure 7.2.1*. Results suggested that this type of sampling grid provides some advantages, however, these advantages cannot be traced back to the sampling approach alone

because a dummy head without torso was simulated and measurement noise is naturally not present in simulated HRTFs.

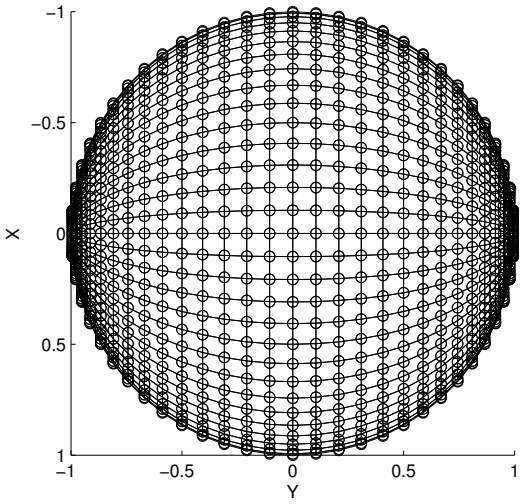
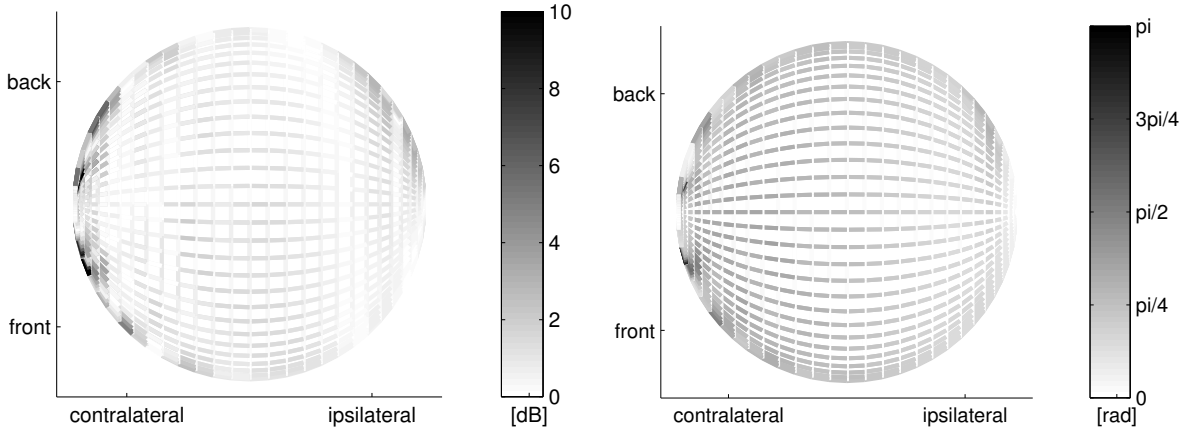


Figure 7.2.1: Sampling grid used for BEM simulation

An initial motivation for this thesis was the assumption of an almost rotationally symmetric HRTF phase response with respect to the interaural axis. Considering the propagation-time dependency of the phase, it is obvious that phase variations occur mainly in lateral directions, as opposed to polar plane directions. An example of the absolute discrete phase gradients of the simulated Neumann head HRTFs is depicted in *figure 7.2.2(b)*. The colours of the lines specify the amount of phase variation along the respective paths. Clearly, the phase variation is very low within sagittal planes and much higher for lateral changes.

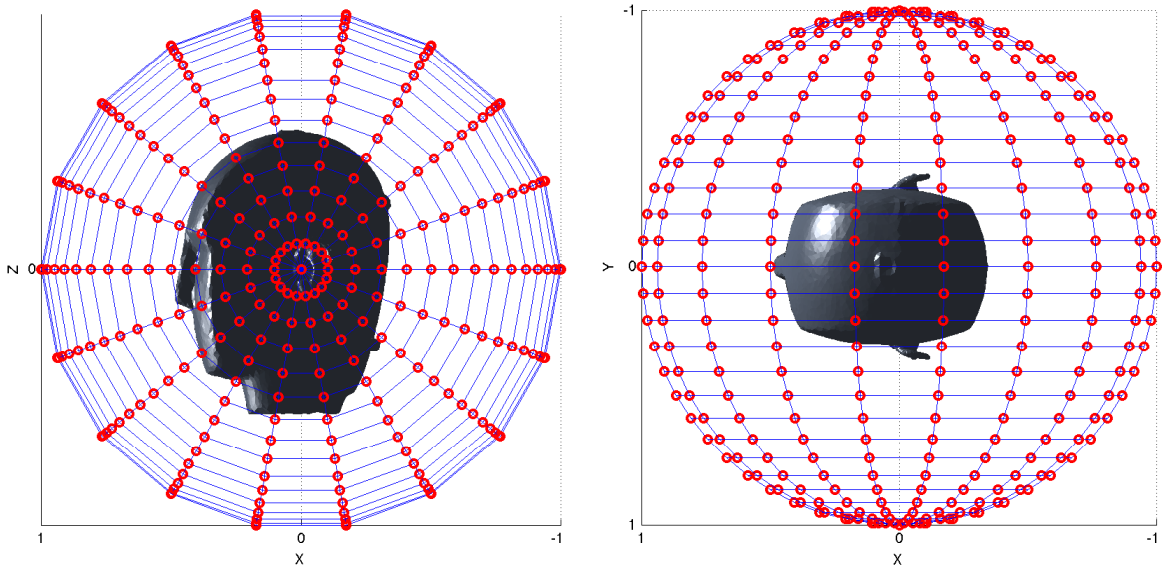


(a) Magnitude (b) Phase

Figure 7.2.2: Absolute discrete gradients for $f = 4\mathrm{kHz}$, view from above the head

Phase and magnitude are naturally measured for the same positions. The magnitude responses must show variations in sagittal planes due to the well-known pinna cues required for cone

of confusion disambiguation. The absolute discrete gradients of the corresponding magnitude distribution are depicted in *figure 7.2.2(a)*. Although the magnitude distribution in sagittal planes is clearly more complicated than the phase distribution, its variations also seem to be higher for lateral changes.



(a) View from the left (b) View from above

Figure 7.2.3: Proposed sampling scheme

A sampling scheme using high density in lateral direction and lower density in polar plane direction, as depicted in *figure 7.2.3*, might hence be beneficial. A standard HRTF database that uses a comparable spatial sampling approach is the KEMAR database. Furthermore, this sampling scheme is well-adapted to the requirements of the spherical phase unwrapping algorithm and might also facilitate a further reduction of the required spherical harmonics by spherical harmonic band limitation with regard to the degree m. A further optimization of the sampling scheme can be achieved based on the approach to finding optimized sampling grids introduced in [Köß11]. It uses a given dense discrete grid and iteratively determines a subset of points with low condition number for optimal DSHT.

Polar gaps. The HRTF data used in this thesis are measured or simulated on a full-sphere sampling grid, which allows for convenient matrix inversion using the DSHT. Since HRTFs are typically measured on a partial sphere with a polar gap in the region below the head, the method presented in this thesis is not directly applicable to standard HRTFs. In order to avoid regularization problems and thus an increase of necessary spherical harmonics, it is suggested to mirror filters from positions in the upper hemisphere to the polar gap region in the lower hemisphere. This might even enforce the rotational symmetry of magnitude and phase responses in space, since the influence of the torso is partly neglected.

Psychoacoustic verfication of results. The analysis of the results has so far only been performed analytically using psychoacoustic error measures. In order to verify the analytic results, a psychoacoustic experiment is necessary. This should ideally include a detailed investigation of the minimal sampling grid density required for an accurate interpolation of magnitude and spherically unwrapped phase responses via the DSHT. The considerations about sampling strategies and possible adaptations of spherical phase unwrapping for coarse sampling grids might also be rewarding questions to evaluate on a psychoacoustic basis.

Bibliography

[AAD01] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc*, vol. 49, pp. 472–479, 2001.

[BDH96] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, pp. 469–483, 1996.

[BNK05] S. Busson, R. Nicol, and B. Katz, "Subjective investigations of the interaural time difference in the horizontal plane," in *Audio Engineering Society Convention 118*, 2005.

[Bus06] S. Busson, "Individualisation d'indices acoustiques pour la synthèse binaurale," Ph.D. dissertation, Académie d'Aix-Marseille Université de la mediterranée Aix-Marseille II, 2006.

[CMM+02] M. Constantini, F. Malvarossa, F. Minati, L. Pietranera, and G. Milito, "A three-dimensional phase unwrapping algorithm for processing of multitemporal sar interferometric measurements," in *Geoscience and Remote Sensing Symposium*, 2002.

[CR99] M. Costantini and P. Rosen, "A generalized phase unwrapping approach for sparse data," in *Geoscience and Remote Sensing Symposium*, 1999.

[Del34] B. Delaunay, "Sur la sphère vide," *Bulletin de l'académie des sciences de l'URSS*, vol. 1, pp. 793–800, 1934.

[DM98] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058, 1998.

[DZG04] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of hrtfs," in *In IEEE ICASSP'04*, 2004, pp. 45–48.

[EAT98] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 1998.

[FLW99] D. J. K. Frederic L. Wightman, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.*, vol. 105, pp. 2841–2853, 1999.

[GK07] R. Greff and B. Katz, "Round robin comparison of hrtf simulation results: Preliminary results." in *roc. 123rd Convention of the Audio Engineering Society*, 2007.

[GODZ10] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 370–386, 2010.

[GZW88] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: Two-dimensional phase unwrapping," *Radio Science*, vol. 23, pp. 713–720, 1988.

[Khu77] G. Khun, "Model for the interaural time difference in the azimuthal plane," *J. Acoust. Soc. Am.*, vol. 62, pp. 157–167, 1977.

[KIC99] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.

[KW92] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[Köß11] T. Kößler, "Kugelkappenförmige mikrofonanordnung zur 3-dimensionalen erfassung des schallfeldes," Master's thesis, Institut für Elektronische Musik und Akustik, Unversität für Musik und darstellende Kunst, Graz, 2011.

[LJ99] V. Larcher and J.-M. Jot, "Techniques d'interpolation de filtres audionumériques: Application à la reproduction spatiale des sons sur écouteurs," in *Congrès Français d'Acoustique*, 1999.

[Mat06] G. R. V. Matias, "Radar interferometry: 2d phase unwrapping via graph cuts," Master's thesis, Lisbon Technical University, 2006.

[MM02] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, vol. 111, pp. 2219–2236, 2002.

[MPC05] P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *J. Audio Eng. Soc*, vol. 53, no. 10, pp. 919–929, 2005.

[MPO+00] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, "The interaural time difference in binaural synthesis," in *Audio Engineering Society Convention 108*, 2000.

[NCNW10] K.-V. Nguyen, T. Carpentier, M. Noisternig, and O. Warusfel, "Calculation of head related transfer functions in the proximity region using spherical harmonics decompositions: Comparison with measurements and evaluation," in *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, 2010.

[NKA08] J. Nam, M. A. Kolar, and J. S. Abel, "On the minimum-phase nature of head-related transfer functions," in *Audio Engineering Society Convention 125*, 2008.

[NSMH03] M. Noisternig, A. Sontacchi, T. Musil, and R. Höldrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, 6 2003.

[NZK11] M. Noisternig, F. Zotter, and B. Katz, "Reconstructing sound source directivity in virtual acoustic environments," in *Principles and Applications of Spatial Hearing*, Y. Suzuki, D. Brungart, and H. Kato, Eds. World Scientific Publishing Co Pte Ltd, 2011.

[OISI10] M. Otani, Y. Iwaya, Y. Suzuki, and K. Itoh, "Numerical analysis of hrtf spectral characteristics based on sound pressures on a pinna surface," in *Proc. of the 20th International Congress on Acoustics (ICA2010)*, 2010.

[Pom08] H. Pomberger, "Angular and radial directivity control for spherical loudspeakerarrays," Thesis, Institut für Elektronische Musik und Akustik, Kunstuni Graz,Technical University Graz, Graz, A, 2008.

[SHLV99] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc*, vol. 47, pp. 675–705, 1999.

[SI07] J. O. Smith III, *Introduction to Digital Filters with Audio Applications*, J. O. Smith III, Ed. BookSurge Publishing, 2007. [Online]. Available: http://www.dsprelated.com/dspbooks/filters/

[Spa95] U. Spagnolini, "2-d phase unwrapping and instantaneous frequency estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 579–589, 1995.

[TM08] D. Toledo and H. Møller, "Audibility of high q-factor all-pass components in head-related transfer functions," in *Audio Engineering Society Convention 125*, 10 2008.

[Wes] [Online]. Available: http://193.166.3.2/pub/sci/audio/jwest-thesis/

[Wik] [Online]. Available: http://en.wikipedia.org/wiki/Delaunay_triangulation

[Wil99] E. G. Williams, *Fourier Acoustics*. Academic Press, 1999.

[ZDG09] D. Zotkin, R. Duraiswami, and N. Gumerov, "Regularized hrtf fitting using spherical harmonics," in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on*, 10 2009, pp. 257 –260.

[Zot09] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," Ph.D. dissertation, Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, 2009.

Chapter 8

Annex

8.1 Minimum Phase Calculation

Phase response. Minimum phase systems have their poles and zeros inside the unit circle and are therefore stable and causal and have causal and stable inverses. Every symmetric magnitude spectrum of a real-valued impulse response $|H[\omega_l]| = |DFT\{h[k]\}|$ has a unique corresponding minimum phase impulse response $h_{\min}$, and hence a unique corresponding minimum phase response $\phi_{\min}$. The minimum phase response and magnitude spectrum are connected via the Hilbert transform $\mathcal{H}$ [Bus06]

$$\phi_{\min}[\omega_l] = \Im\{\mathcal{H}\{-\ln(|H[\omega_l]|)\}\},$$

where $\Im$ denotes the imaginary part.
Another approach uses the real cepstrum $h_{\text{rceps}}[k] = IDFT\{\ln(|H[\omega_l]|)\}$, where the anticausal part needs to be flipped around time zero so it adds to the causal part [SI07]

$$h_{\text{rceps,causal}}[k] = \begin{cases} h_{\text{rceps}}[k] + h_{\text{rceps}}[-k] & \text{for } 0 < k \leq \frac{\mathrm{K}}{2}, \\ h_{\text{rceps}}[k] & \text{for } k = 0,\ \frac{\mathrm{K}}{2} + 1. \end{cases}$$

The minimum phase is then given by

$$\phi_{\min}[\omega_l] = \angle\ \mathrm{e}^{DFT\{h_{\text{rceps,causal}}[k]\}}.$$

Impulse response. The related minimum phase impulse response $h_{\min}[k]$ can either simply be calculated using $|H[\omega_l]|$ and $\phi_{\min}[\omega_l]$

$$h_{\min}[k] = IDFT\{|H[\omega_l]|\ \mathrm{e}^{\mathrm{j}\phi_{\min}[\omega_l]}\},$$

or directly from the causal real cepstrum $h_{\mathrm{rceps,causal}}[k]$

$$h_{\mathrm{min}}[k] = IDFT\{\mathrm{e}^{DFT\{h_{\mathrm{rceps,causal}}[k]\}}\}.$$

8.2 Onset Delay Extraction

There is a variety of approaches for the calculation of the temporal onset delay. In the context of HRIRs, it is quite usual to use the difference between the right and left onset delays, i.e. the ITD [BNK05, MPO+00, KW92]. In this thesis, the absolute delays contained in the right and left HRIRs are calculated. An accurate and widely applied method for ITD detection is the *maximum IACC*, where the maximum of the interaural cross-correlation function is calculated

$$ITD = \mathrm{argmax}\ IDFT\{H_{\mathrm{r}}[\omega_l] \cdot H_{\mathrm{l}}[\omega_l]^*\}.$$

$H_{\mathrm{r}}[\omega_l]$ and $H_{\mathrm{l}}[\omega_l]$ denote the complex spectra of the right and left HRIRs, respectively. This approach can easily be adapted for the estimation of the absolute delay τ, using the spectra of an impulse reponse ($H[\omega_l]$) and its minimum phase version $H_{\mathrm{min}}[\omega_l]$ for calculating the cross-correlation

$$\begin{aligned} \tau[\mathrm{samples}] &= \mathrm{argmax}\ IDFT\{H[\omega_l] \cdot H_{\mathrm{min}}[\omega_l]^*\}, \\ \tau[\mathrm{s}] &= \mathrm{argmax}\ IDFT\{H[\omega_l] \cdot H_{\mathrm{min}}[\omega_l]^*\}/\mathrm{f_s}, \end{aligned}$$

where $\mathrm{f_s}$ denotes the sampling rate. Since the minimum phase version of the impulse response always has its peak at the first sample, the onset delay τ results from the cross-correlation.

8.3 Onset Delay Simulation

In the field of binaural synthesis, a variety of mathematical models have been proposed for the simulation of the ITD [Khu77, LJ99, SHLV99, AAD01]. In order to simulate the onset delay, a spherical head model is used. The following derivations are performed in 2D, i.e. using a circle instead of a sphere. Nevertheless, the mathematical relations hold for the spherical case as well.

First case: Direct path. Considering a model as depicted in *figure 8.3.1*, with the head radius **r** , the measurement radius **R**, the position of the ear **E**, and the measurement position **P**, the length **S** of the acoustic path between the measurement position and the ear can directly be calculated as

$$\mathbf{S} = \|\vec{\mathbf{E}} - \vec{\mathbf{P}}\|,$$

provided that the angle $\psi \geq 90°$, which leads to the *constraint for the direct path*:

$$\mathbf{r}^2 + \mathbf{S}^2 \leq \mathbf{R}^2.$$

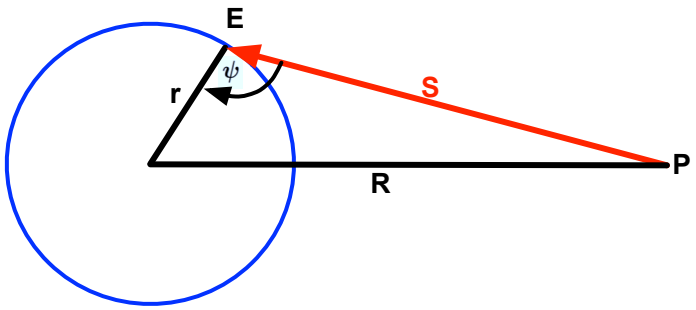


Figure 8.3.1: Spherical model: direct path

Second case: Bent path. If $\psi \leq 90°$, the ear position is placed at the averted side of the sphere, so the path needs to be split into two parts (see *figure 8.3.2*). The *constraint for the bent path* is

$$\mathbf{r}^2 + \mathbf{D}^2 > \mathbf{R}^2$$

The first part $\mathbf{S}_1$ represents the direct connection between the measurement position **P** and the tangent point **T**. It can be calculated using the Pythagorean theorem:

$$\mathbf{S}_1 = \sqrt{\mathbf{R}^2 - \mathbf{r}^2}.$$

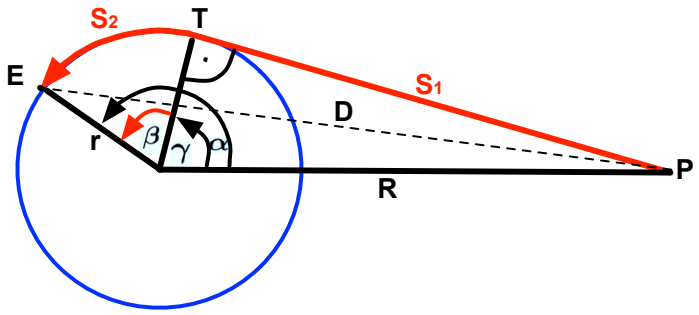


Figure 8.3.2: Spherical model: bent path

The second part S_2 is given by

$$\mathsf{S}_2 = \mathsf{r}\ \boldsymbol{\beta},$$

where

$$\boldsymbol{\beta} = \boldsymbol{\alpha} - \boldsymbol{\gamma} = \arccos\left(\frac{\vec{\mathsf{E}} \circ \vec{\mathsf{P}}}{\mathsf{r}\ \mathsf{R}}\right) - \arccos\left(\frac{\mathsf{r}}{\mathsf{R}}\right).$$

The total length of the acoustic path writes as

$$\mathsf{S} = \mathsf{S}_1 + \mathsf{S}_2.$$

Calculation of delay. The onset delay τ can now be calculated from the length of the acoustic path S

$$\begin{aligned} \tau &= \frac{\mathsf{S}[\mathrm{m}]}{\mathsf{C}[\mathrm{m/s}]} \text{ in } [\mathrm{s}], \\ \tau &= \frac{\mathsf{S}[\mathrm{m}]}{\mathsf{C}[\mathrm{m/s}]}\ \mathrm{f_s} \text{ in } [\mathrm{samples}], \end{aligned}$$

where C and $\mathrm{f_s}$ denote the sound velocity and the sampling rate, respectively.