Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Radioelectronics

SOURCE LOCALIZATION BY VIRTUAL ACOUSTIC REALITY:
Differential Head-Related Transfer Function as a One-Channel Positioning Method

by

Dominik STOREK

Doctoral thesis submitted to
the Faculty of Electrical Engineering, Czech Technical University in Prague,
in partial fulfilment of the requirements for the degree of Doctor.

PhD programme: Electrical Engineering and Information Technology
Branch of Study: Acoustics

Prague, October 2016

Thesis Supervisor:

PETR MARSALEK
Department of Pathological Physiology
First Medical Faculty
Charles University in Prague
U Nemocnice 5
128 53, Prague 2
Czech Republic

Thesis Co-Supervisor:

FRANTISEK RUND
Department of Radioelectronics
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2
166 27, Prague 6
Czech Republic

Abstract

This thesis presents a new algorithm for virtual sound source positioning for the purposes of rendering acoustic components of headphone-based virtual reality. The one-channel positioning method using Differential Head-Related Transfer Function (DHRTF) is introduced. The DHRTF method utilizes localization cues (time and intensity differences) extracted from HRTF pairs for particular positions. In contrast to the usual two-channel filtering by a HRTF pair, only one channel is processed when the DHRTF algorithm is applied. This results in one channel is being delayed and attenuated with respect to the other one, which contains the unprocessed original sound. Therefore, the final positioned sound has the same time and level differences as when filtered by the HRTF pair, however, with different signals in each channel.

An introduction to the field of virtual positioning and contemporary state-of-the-art presents the first chapter. The theoretical concept of the DHRTF positioning method is presented and demonstrated with model examples in the second chapter. Specific features, performance, and efficiency of the DHRTF algorithm are examined and compared to the attributes of other positioning methods (namely amplitude panning and HRTF filtering) in the last chapter.

The highlights of the results delineate remarkably good spatial depth, reduction of processing requirements, and low affection of natural sound timbre. Possible applications of the DHRTF method are in assistive systems for visually impaired, in computer games audio, and in enhancing spatialization of musical or film mixing on digital audio workstations.

Keywords:

Virtual sound source positioning, positioning algorithm, positioning method, virtual auditory space, head-related transfer function, one-channel processing, differential head-related transfer funciton.

Abstrakt

Tato práce se zabývá algoritmy virtuálního polohování zdroje zvuku pro účely generování akustické složky virtuální reality s poslechem přes sluchátka. Je zde představena jednokanálová polohovací metoda založená na aplikaci diferenciální přenosové funkce hlavy (Differential Head-Related Transfer Function - DHRTF). Zmíněná DHRTF metoda využívá časové a intenzitní rozdíly extrahované z páru HRTF pro příslušnou pozici. Oproti běžné dvoukanálové filtraci při použití páru HRTF se v algoritmu využívajícím DHRTF zpracovává pouze jeden kanál. Tento kanál je vždy časově zpožděn a zatlumen oproti druhému, který obsahuje původní nezměněný signál. Výsledný polohovaný zvuk tak disponuje stejnými časovými a intenzitními rozdíly mezi kanály jako v případě filtrace pomocí HRTF, avšak s rozdílným signálovým obsahem.

První kapitola uvede do problematiky virtuálního polohování zdroje zvuku a se současného stavu výzkumu. Teoretický koncept polohovací metody založené na DHRTF je představen a demonstrován na jednoduchém modelu v kapitole druhé. Specifickými vlastnostmi této metody a její efektivitou ve srovnání s ostatními polohovacími metodami (panorama, HRTF) se zabývá kapitola poslední.

Dosažené výsledky popisují vlastnosti představované DHRTF metody. Konkrétně se jedná o dobrou hloubku vnímaného prostoru v polohovaném zvuku, redukci výpočetních nároků a nízký vliv na přirozenou barvu vnímaného zvuku. Možné aplikace DHRTF polohovacího algoritmu jsou například asistivní systémy pro nevidomé, použití v herním *audio engine*, či dosažení větší prostorovosti při mixu hudebního nebo filmového díla bez nechtěné ekvalizace.

Klíčová slova:

Virtuální polohování zdroje zvuku, polohovací algoritmus, polohovací metoda, virtuální akustický prostor, head-related transfer function, jednokanálové zpracování, diferenciální přenosová funkce hlavy.

Declaration

I hereby declare that I worked out the presented thesis independently and I quoted all used sources of information in accord with Methodical instructions about ethical principles for writing academic thesis.

....

Dominik Štorek

Acknowledgements

This thesis would not have been created without a contribution of particular individuals, who helped me during my research. Hereby I would like to express my gratitude to my supervisor, prof. RNDr. Petr Maršálek, Ph.D. for his support and guidance. Great thanks also belong to doc. Ing. František Kadlec, CSc., who has been my supervisor until his retirement at the beginning of my PhD studies.

I would also like to thank to Ing. František Rund, Ph.D. for his valuable comments regarding my work and fruitful cooperation not only within our grant assignments. I also very appreciate moral and vocational contribution of Dr. Ing. Libor Husník and his initial key role in arrangement of my abroad internship. At this point, I would also like to thank to prof. Dr. Paul Wai-Fung Poon for his kindness and personal handling during the internship program at Dept. of Physiology, Medical College, National Cheng Kung University, Tainan City, Taiwan.

Special thanks go to the colleagues from the Department of Radioelectronics, who have been valuable partners for discussions of both professional and personal issues, namely Ing. Filip Fikejz, Ing. Václav Vencovský, Ph.D., Ing. Jaroslav Bouše, and Ing. Lukáš Krasula from CTU in Prague and also Da-Wei 'Dannel' Shen and I-Chieh 'Rypia' Huang from the department of Physiology at National Cheng Kung University, Tainan, Taiwan. My thanks also belong to Craig E. Smith for English language corrections. Special thanks are indeed addressed to T. Reks and S. Uperjaarda for inspiration.

My research has been partially supported by the Ministry of Education, Youth, and Sport of the Czech Republic under research program MSM 6840770014, and by the Grant Agency of the Czech Technical University in Prague (SGS) as project *Using of the Sound Sources Virtual Positioning Methods for Three-dimensional Orientation of the Visually Impaired* (No. SGS10/082/OHK3/1T/13), project *Personal Navigation System using Artificial Audio Signals for Assistive Technology* (No. SGS11/159/OHK3/3T/13), and finally project *Psychoacustical measurements, auditory models and their application* (No. SGS14/204/OHK3/3T/13).

I would also like to gratefully thank to my family members, for their infinite patience and care not only during my PhD study. Finally, my greatest thanks go to my love Lucie for her support and care, however, primarily for her being. And also for her delicious cookies she made for my final PhD exam and thesis defense.

Dedication

To Lucinka

Contents

Abstract (EN) **iii**

Abstract (CZ) **iv**

List of Figures **xii**

List of Tables **xv**

Abbreviations **xvi**

1 Introduction **1**

1.1 Motivation 1

1.2 Problem Statement 2

1.3 Related Work/Previous Results 2

1.4 Goals of the Thesis 3

1.5 Structure of the Thesis 3

2 State-of-the-Art **5**

2.1 Principles of Sound Source Localization 5

2.2 Basics of Virtual Sound Source Positioning 10

2.2.1 Amplitude Panning 12

2.2.2 Time Shift 13

2.3 Introduction to Head-Related Transfer Function 15

2.4 Virtual Positioning by the HRTF 18

2.4.1 Static Virtual Positioning 18

2.4.2 Dynamic Virtual Positioning 20

2.4.3 Environmental Reverberation 22

2.5 Attributes of HRTF set and HRTF acquisition 23

2.5.1 Measuring of HRTF 23

2.5.2 Modeling of HRTF 27

2.5.3 Representation of a HRTF set 28

2.6 Quality Assessment 32

2.6.1 Measuring Methods 37

3 Proposed Method: Positioning by Differential HRTF 39
3.1 Method Background . . . 39
3.2 Differential Head-Related Transfer Function . . . 41
3.2.1 One-dimensional Demonstration . . . 43
3.2.2 Two-dimensional Demonstration . . . 44
3.3 Virtual Positioning by the DHRTF . . . 47
3.4 DHRTF set from a real measured HRTF . . . 49
3.5 Issues in positioning by Differential HRTF . . . 50

4 Specific Aspects of DHRTF Method 53
4.1 Comparison of Positioning Methods . . . 54
4.1.1 Objective Comparison of Positioning Methods . . . 54
4.1.1.1 Position-dependent Gain . . . 54
4.1.1.2 Position-Dependent Transfer Function . . . 56
4.1.1.3 Time Domain Aspects . . . 59
4.1.2 Subjective Comparison of Positioning Methods . . . 60
4.1.2.1 Listening Test: Positioning Methods Comparison . . . 61
4.1.2.2 Stimuli Description . . . 61
4.1.3 Results . . . 62
4.1.4 Discussion . . . 66
4.2 Implementation of Moving Source . . . 67
4.2.1 Dynamic Positioning with Linear Convolution . . . 67
4.2.2 Dynamic Positioning with Circular Convolution . . . 68
4.2.3 Employing Overlap-Add Method . . . 69
4.2.4 Effect of Window Length . . . 70
4.2.5 Analysis of Dynamic Positioning . . . 72
4.2.6 Listening Tests: Dynamic Positioning . . . 74
4.2.7 Results . . . 75
4.2.8 Discussion . . . 77
4.3 Perception Artifacts in the DHRTF Method . . . 78
4.3.1 Negative ILD and its Impact to DHRTF Positioning . . . 78
4.3.2 Methods of Artifact Reduction . . . 79
4.3.2.1 Spectral Amplitude Limiting . . . 81
4.3.2.2 Low-pass Filtering of Spectral Amplitude . . . 81
4.3.2.3 Moving Average Smoothing of Spectral Amplitude . . . 82
4.3.3 Artifact Reduction in Time and Frequency Domain . . . 83
4.3.4 Listening Tests: Artifact Reduction . . . 85
4.3.4.1 Experiment A: Artifacts in Half-Plane Range . . . 85
4.3.4.2 Experiment B: Focus on Artifacts-Related Positions . . . 87
4.3.5 Objective Assessment . . . 88
4.3.6 Results . . . 88
4.3.7 Discussion . . . 90

5 Conclusions 93
5.1 Summary 93
5.2 Contributions of the Thesis 95
5.3 Future Work 95

Bibliography 97

Publications of the Author 107

A Appendix 111

List of Figures

2.1 Coordinate system 6
2.2 Localization principles 7
2.3 Shoulder reflections 8
2.4 Pinna reflections 9
2.5 Listening systems 11
2.6 Crosstalk in loudspeaker listening 11
2.7 ITD geometry 13
2.8 Evaluation of ITD approximation 14
2.9 Source-listener system 15
2.10 HRIR and HRTF 16
2.11 Static FIR filter structure 18
2.12 Static positioning 19
2.13 Dynamic positioning 21
2.14 Dynamic FIR filter structure 21
2.15 HRTF processing scheme 22
2.16 Environmental reverberance 23
2.17 HRTF measurement settings 24
2.18 Example of HRTF measuring points 25
2.19 Positioning of a stimulus 27
2.20 Model of the HRTF 28
2.21 HRIR for the horizontal plane 29
2.22 HRIR for the median plane 30
2.23 HRTF for the horizontal plane - heat map 31
2.24 HRTF for the median plane - heat map 31
2.25 Rendering simple VAS 33
2.26 Horizontal plane JND 34
2.27 Lateralization vs. Externalization 35
2.28 Cone of confusion 36
2.29 Front-back confusion 36

3.1 HRTF versus DHRTF principle 40
3.2 DHRTF derived from the HRTF pair 42
3.3 Obtaining DHRTF 43

3.4 1-D DHRTF demonstration 44
3.5 Simple HRIR set demonstration 45
3.6 Simple DHRIR set demonstration 46
3.7 ITD from HRTF set 46
3.8 Simple DHRTF set demonstration 47
3.9 HRTF vs DHRTF processing 48
3.10 DHRTF processing scheme 49
3.11 Heat map of DHRTF magnitude 50
3.12 Origin of the DHRTF artifacts 51
3.13 Negative ILD in the DHRTF 52

4.1 Gains of AP, HRTF and DHRTF 55
4.2 AP heat map 56
4.3 HRTF heat map 57
4.4 DHRTF heat map 58
4.5 ILD azimuth functions 59
4.6 Positioning methods on a sine wave 60
4.7 Spatial Impression rating 63
4.8 Coloration rating 64
4.9 Loudness rating 64
4.10 Preferences based on content 66
4.11 Linear convolution-based positioning 68
4.12 Circular convolution-based positioning 69
4.13 One-block demonstration of the processing 70
4.14 Overlapping of segments 71
4.15 Discrete HRTF positions 72
4.16 Switching FIR impulse responses 73
4.17 Convolution of a segment 74
4.18 Application area 76
4.19 Circular vs linear convolution preference 76
4.20 Histogram NILD 79
4.21 Artifacts occurrence 80
4.22 Elimination of artifacts 81
4.23 DHRTF adjustment 82
4.24 dHRIR pre-processing 83
4.25 An original dHRIR set 84
4.26 A processed dHRIR set 84
4.27 Results of Experiment A 90
4.28 Results of Experiment B 91
4.29 Results for Objective assessment 91

List of Tables

4.1 Rating keys . . . 62
4.2 Positioning methods summary . . . 65
4.3 Results of dynamic listening tests . . . 75
4.4 Results of Experiment A . . . 89
4.5 Results of Experiment B . . . 89

Symbols and Abbreviations

Variables & Symbols

φ	angle in horizontal plane
ϑ	angle in median plane
$\Delta\varphi$	sample step in horizontal plane
$\Delta\vartheta$	sample step in median plane
ω	angular frequency of continuous spectral domain
ω_r	angular frequency of moving sound source
ψ	phase vector
t	continuous time
n	discrete time
k	discrete time shift
τ	discrete time expressing variance of parameters
Ω	discrete spectral domain
j	complex unit
p	sound pressure
f_s	sampling frequency
w	window weighting function
g	channel gain
y	output signal (positioned signal)
Y	output spectrum (positioned signal)
x_m	monaural source (original non-positioned signal)
X_m	monaural source spectrum

Mathematical Terminology & Operators

$\boldsymbol{A}$	Matrix $\boldsymbol{A}$
$\boldsymbol{A}_{m,n}$	Element of matrix $\boldsymbol{A}$ on m^{th} row to n^{th} column
$x[n]$	Vector of elements n
$(x[n])_{a:b}$	Selection of a sequence of vector $x[n]$ from a^{th} element to b^{th} element
$\circledast$	Linear Convolution
$\otimes$	Circular Convolution
$\lvert H[z] \rvert$	Module of transfer function $H[z]$
$arg\{H[z]\}$	Phase of transfer function $H[z]$
FT	Fourier transform
FT^{-1}	Inverse Fourier transform
$\lceil \cdot \rceil$	Round up to an integer (*ceil*)
$\lfloor \cdot \rfloor$	Round down to an integer (*floor*)
$a \in (b, c)$	a lies in set specified by interval from b to c
$a \in \{b, c\}$	a is one of the objects of a set comprised of b and c

Miscellaneous Indices

ξ	Correspondence to either *left* or *right* channel.
L	Correspondence to *left* channel.
R	Correspondence to *right* channel.
c	Correspondence to *contra-lateral* channel.
i	Correspondence to *ipsi-lateral* channel.
φ	Position in *horizontal* plane
ϑ	Position in *median* plane
HT	Head-Tracking
NHT	Non-Head-Tracking
ART	Processing artifacts
JIT	Positioning jitter

Miscellaneous Abbreviations

ALE	Average Localization Error
AP	Amplitude Panning
BEM	Boundary Element Method
DHRIR	Differential Head-Related Impulse Response
DHRTF	Differential Head-Related Transfer Function
FBC	Front-Back Confusion
HHS	Human Hearing System
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
HT	Head-Tracking
ILD	Interaural Level Difference
IPD	Interaural Phase Difference
ITD	Interaural Time Difference
ITF	Interaural Transfer Function
JND	Just Noticeable Difference
LTI	Linear Time-invariant
MA	Moving Average
MAA	Minimum Audible Angle
ODG	Objective Difference Grade
RIR	Room Impulse Response
RMS	Root Mean Square
SL	Sine Law
VAS	Virtual Acoustic Space
VR	Virtual Reality

Chapter 1

Introduction

This doctoral thesis deals with algorithms for virtual sound source positioning for the purpose of rendering the acoustic components of virtual reality. This chapter briefly introduces the motivation for writing this work, defines specific challenges to be discussed and encountered, and summarizes the background. Furthermore, the structure of the thesis concept is introduced and main contributions of the thesis are emphasized.

1.1 Motivation

Human hearing is a unique system of two acoustic receivers, that allows the higher hearing centers not only to extract and distinguish the content of acoustic information from the surrounding environment, but also evaluate positions of the sound source and deliver information about the space arrangement of the surrounding sound-active objects. This ability is very important for orientation in space and historically necessary for immediate reaction within crisis situations.

The most important sense for orientation in space for humans is obviously the sense of sight. Spatial hearing enhances the sight's ability of spatial orientation and supports it by additional information on different basis. The importance of the hearing component increases rapidly for insufficient sight conditions (dark, fog, smoke, etc.). In these situations, hearing is the only sense which most humans can rely on.

The ability to distinguish particular sound source position is based on differences between the acoustic signal in left and right ear channels and also the specific monaural character of the sound. Entrances of both ear channels are spatially separated, thus the conditions for receiving the propagated sound wave are different. Arrival time and amplitude levels of sound signals at positions of the ear channel entrances vary according to the position of sound source. The body structure in its full size also constitutes an obstacle or reflector for the sound waves of different wavelengths. Therefore, specific features like reflection and diffraction occur on the path from the sound source to the listener's eardrum. Such specific character of the sound results in particular perceived position of the source, represented by unique combinations of level, time, and spectral parameters of the received signal. These parameters are dependent on many factors, as described in the following chapters.

The main goal of virtual sound source positioning is literally to *hoodwink* the brain evaluation process of the source position and create an illusion of sound coming from the desired location. Therefore, the listener perceives not only the acoustic information of the source itself, but also its virtual position. This approach involves a process of altering parameters of the physical signal destined for either ear channel in such way that the final changes do not directly affect the perception of the sound information itself, but extends the original by cues for spatial arrangement for both the source and the listener. These modifications shall be in accordance with phenomena occurring in real listening environment.

Virtual sound source positioning is a process of spatial separation of the sources in rendered audio scenes. For its performance either a set of loudspeakers or a headphone-based system is used to produce the virtual scene [1]. In listening systems based on a physical arrangement of loudspeakers (e.g. cinema sets, N.1 home sets), the signal is distributed among multiple channels in order to achieve the spatial illusion for the listener. However, the headphone-based systems are generally limited to only two channels.

The interaction between the sound wave and the listener's body can be described by *Head Related Transfer Function* (HRTF) [2]. Every individual has a unique HRTF, consequently an HRTF based on a prototypical listener's head can be used [3, 4, 5]. In virtual positioning it is necessary to simulate these acoustical features in both ear channels [6, 7, 8], which leads to the application of the HRTF as a channel transfer function to either channel in the binaural pair.

1.2 Problem Statement

A deep exploration of the field of virtual sound source positioning has revealed multiple issues. Introduction to this field is described in chapter 2, beginning on page 5. One of the issues is compression of spatial information used in the virtual positioning and reduction of processing complexity while preserving maximal lifelike spatial perception. Recently, several algorithms of various complexity, implementation requirements, and resulting spatial fidelity are known. These algorithms are utilized according to the requirements of a specific application. Demands for an algorithm that preserves maximum spatial information while reducing its implementation requirements have arose.

1.3 Related Work/Previous Results

In prior work, the author participated in research dealing with the possibilities of constructing assisstive sonification devices for the visually impaired. In this field, the requirement for a sufficient ratio between simplicity and effectiveness is crucial. Previous results regarding virtual positioning, which have been gained in this period, are briefly pointed throughout the text, mostly as a reference to related articles. The list of all relevant articles completed up to the date of finishing this thesis is available at the end of the work.

1.4 Goals of the Thesis

The primary goal of this thesis is to develop a new positioning algorithm that is designed to reduce computational complexity while rendering the positioned sound. A synopse of the particular goals is as follows:

1. Design of a new positioning algorithm, which enables reduction of the virtual positioning process from the usual double-channel processing to only one-channel processing. This algorithm is based on utilizing *Differential Head-Related Transfer Function* (DHRTF).

2. Introduction of principles of the method and acquisition of the DHRTF from an existing pair of HRTFs. Demonstration of localization cues extraction are included.

3. Implementation of the DHRTF algorithm and its theoretical comparison to common positioning algorithms for both static and dynamic virtual positioning.

4. Design, realization, and evaluation of multiple specific listening tests in order to compare the DHRTF algorithm to the common positioning methods in terms of the quality of spatial perception and character of the final processed sound.

5. Exploration of the artifacts which may occur in the virtual positioned sound as a result of specific processing features and detailed design of methods for their elimination verified by listening tests.

6. Organizing the state-of-the-art of headphone-based virtual sound source positioning into a (hopefully) neatly integrated survey introduced at the beginning of the thesis.

1.5 Structure of the Thesis

The thesis is organized into 5 main chapters. It both introduces the research to the reader and determines its current status in virtual positioning. The chapters are organized as follows:

1. *Introduction*: Describes the motivation behind the research efforts and its goals. There is also a list of particular contributions of this doctoral thesis.

2. *State-of-the-Art*: This chapter introduces the reader to the necessary theoretical background and surveys in current state-of-the-art.

3. *Proposed Method*: Introduces and explains the essential ideas of the proposed positioning method; one-channel positioning by Differential HRTF. This chapter demonstrates the principles and discovers the background of DHRTF acquisition from existing HRTF sets. Issues conjunctive with this processing are also discussed.

4. *Specific Aspects of DHRTF Method*: This part introduces the main results and achievements in particular sections, where each section is devoted to a specific issue. Practical implementation of the positioning algorithm, design, and performance of the listening tests and methods for artifact reduction are presented here. This chapter consists mostly of material published in several papers.

5. *Conclusions*: Summarizes and discusses the results and contributions, suggests possible courses for further research, and concludes the thesis.

Literature used in this doctoral thesis is standardly referred to by a number in square brackets, e.g. [0]. Moreover, work published by the author or work, where the author has participated, is referred to by a prefix *A*, e.g. [A.0]. Each article is marked either IF, RJ, or CO as a reference to either impact factor journal, reviewed journal, or conference contribution, respectively. Separate contents of references are provided at the end of the doctoral thesis in sections *References* and *Publications of the Author*.

Chapter 2

State-of-the-Art

Sound source localization is a process of the hearing sense, which provides the ability to determine the position of the sound source in space without contribution of other senses. This ability is crucial for orientation in the dynamic environment. In normal conditions, hearing significantly supports the visual information. Regarding rendering virtual acoustic reality, it is necessary to understand the mechanisms of sound source localization in space and emulate these phenomena under artificial conditions. This chapter provides a brief overview of the basic principles of spatial sound source localization and its employment in source positioning for virtual acoustic reality. The main principles and phenomena are presented first, then the Head-Related Transfer Function (HRTF) is introduced in more detailed way, since it is a crucial basis for the author's positioning method proposed in this thesis. More information can be found in the references noticed throughout the text.

2.1 Principles of Sound Source Localization

For better understanding, it is necessary to introduce basic terms and features involved in spatial hearing. For a description of spatial localization, a *spherical coordinate system* is used [9] [10]. The origin of the coordinates is considered to be placed at the center (inside) of the listener's head. Three planes can be defined. *Horizontal* plane enables a left-right description, *median* plane enables an up-down description, and *frontal* plane enables a front-back description.

In this system, the sound source position is determined by two angles and distance. Angle φ (azimuth) describes the position in the *horizontal* plane (i.e. left-right movement), angle ϑ (elevation) corresponds to position in the *median* plane (i.e. up-down movement). Particular intervals are defined as $\varphi \in (-180, 180), \vartheta \in (-90, 90)$. This system is demonstrated in Fig. 2.1

When the sound propagates from the source to the listener's ear, the sound wave interacts with the body structure - with torso, head, and pinna. The interaction involves *diffraction* and *reflection*, and finally results in different signal attributes around the listener's body. According to size of each part of the body, different frequency bands are

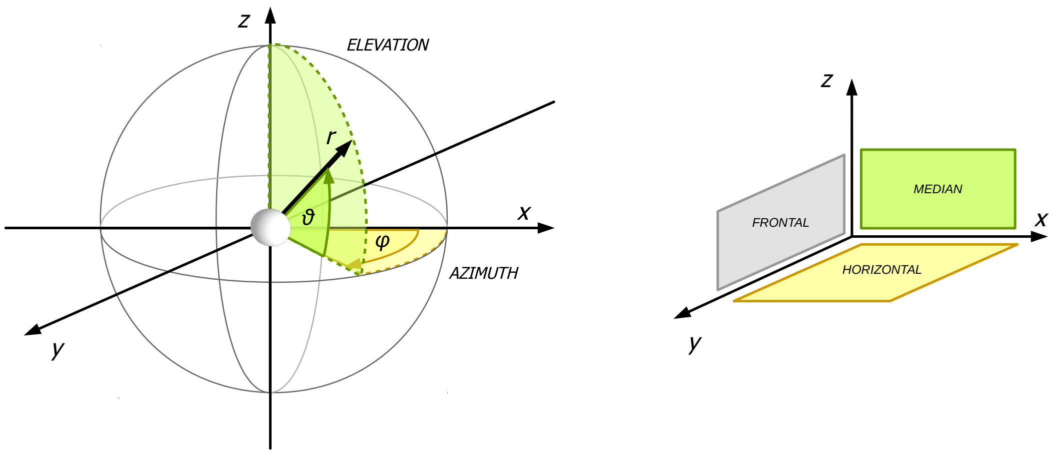

Figure 2.1: Coordinate system. For sound source localization, a spherical coordinate system and three particular planes are used for the description of sound source position.

affected according to its ratio to the wavelength [11], [12].

The *Human hearing system* (HHS) is equipped by two symmetrically placed acoustic sensors in spatially-separated positions [13]. This arrangement allows gathering acoustic information in two different locations, which is the essential factor for the ability of sound source localization [14]. For the purposes of this thesis, the ear located in the same *horizontal* half-plane (closer), defined by x axis, as the sound source, is called *ipsi-lateral* and the ear located in the latter (farther) half-plane is called *contra-lateral* [9]. For further description, the assume spatial arrangement of the listener and the sound source, as demonstrated in Fig. 2.2. The essence of localization ability in the horizontal plane is based on evaluation of tiny differences of the signal in particular ear channels; in amplitude and time of incidence [13]. For a real sound source placed within the horizontal plane (out of the frontal axis), the incident sound wave reaches the (*contra-lateral*) ear with the delay time corresponding to its longer pathway, given by the speed of sound in the air, and with lower intensity. According to the spatial arrangement of the sound source and the listener, two dominant phenomena can be observed [9], [15]:

- *Interaural Time Difference* (ITD)
  In geometrical arrangement, when the source deflects from the front-back x axis, the time of incidence of the sound wave in both ears differs according to the resulting sound propagation geometry. The absolute delay time is determined by the distance difference and speed of sound defining time needed to overcome this distance. The ITD is a function of angle φ and for harmonic signal with wavelength λ, it can be defined as

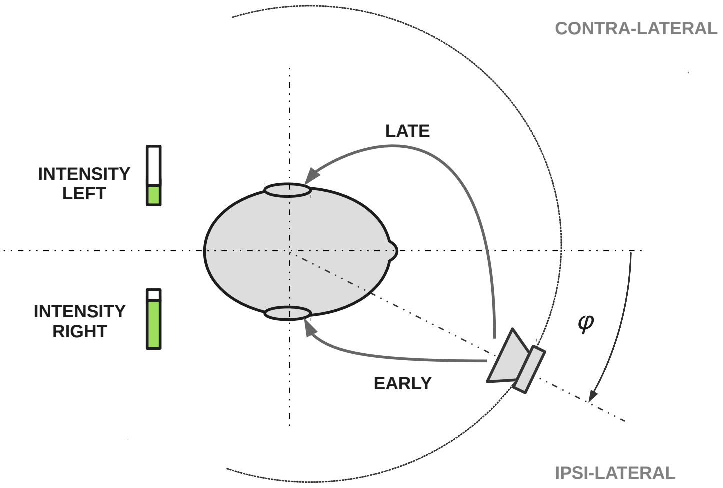

Figure 2.2: Localization principles. Typical time and intensity differences in particular geometric arrangement of the listener and the sound source is demonstrated. The contra-lateral signal reaches the ear later with lower intensity.

$$\mathrm{ITD}(\varphi) = |t_L - t_R| = \frac{|\psi_L(\varphi) - \psi_R(\varphi)|\lambda}{2\pi c_0}, \tag{2.1}$$

where t is the time of incidence indexed by left/right ear and ψ stands for phase. λ refers to wavelength and c_0 to the speed of sound ($\sim$ 340 m $\cdot$ s^{-1}). The magnitude the ITD strongly depends on the head size.

- *Interaural Level Difference* (ILD)
  In geometrical arrangement, when the source deflects from the front-back x axis, the head becomes an obstacle for the sound wave. According to the relation of the wavelength and the head size, diffraction or reflection occur. Therefore the signal reaching the contra-lateral ear is attenuated in comparison to the signal reaching the ipsi-lateral ear in the direct path [16]. The ILD as a function of φ can be defined as

$$\mathrm{ILD}(\varphi) = |L_{p,L}(\varphi) - L_{p,R}(\varphi)| = \left|20 \cdot \log_{10}\left(\frac{p_{ef,L}(\varphi)}{p_{ef,R}(\varphi)}\right)\right|, \tag{2.2}$$

where L_p is the sound pressure level at the entrance of the ear channel and p_{ef} stands for effective value of the sound pressure, indexed by corresponding ear channel. The amount of the ILD varies not only according to the source position, but it is also directly connected to the incident wavelength. The path of the wave is diffracted to

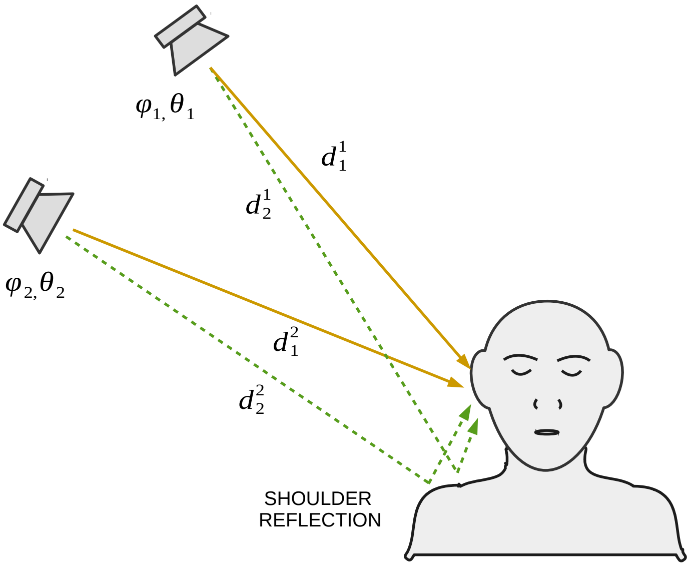

Figure 2.3: Shoulder reflections. Reflections for different source positions are shown. Particular positions φ_1, ϑ_1 and φ_2, ϑ_2 result in different ratios of the direct path d_1 and reflected path d_2.

the contra-lateral ear for lower frequencies (up to app. 1500 Hz [9]), when wavelengths are longer than the head diameter. Higher frequencies are reflected by the head, since the wavelength is shorter than the head size. The ILD on particular frequencies can raise over 50 dB [17].

In natural listening environment, both ITD and ILD are frequency dependent [16], [18]. In binaural hearing, the boundary between low and high frequencies is approximately 1.5 kHz (with respect to anthropometric parameters). The ITD effects occur at lower frequencies and the effects of the ILD are present at higher frequency range [2]. This is determined by the mechanisms of signal coding in the inner ear [9] and in subsequent neurons of the auditory pathway [19, 20, 21, 22].

However, there are different mechanisms for localization in the median plane. For the central position ($\varphi = 0$), when the source is placed on the horizontal x axis, the path for sound coming to both ears is (generally) identical, thus both ITD and ILD do not occur. The elevation ϑ of the sound source measured as an angle within the median plane is perceived primarily due to the propagation of high frequencies and its reflection on the surface of the outer ear (pinna) [23], and reflection from shoulder [24]. These spectral components (named monaural cues or spectral cues [25]) can be observed above approximately $f = 6$ kHz [2] and their character varies according to elevation of the source. In the median plane localization, the importance of body reflection increases [26, 27]. The first phenomenon deals with *shoulder reflection* of the sound [28]. This effect is

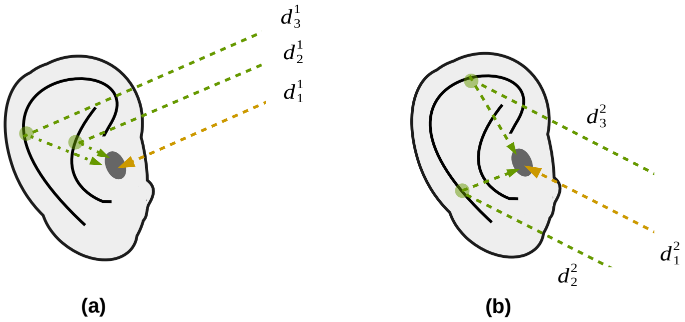

Figure 2.4: Pinna reflections. A scheme of multiple reflections on the pinna surface is depicted. Different path lengths result in a unique comb filter effect for a particular arrangement.

demonstrated in Fig. 2.3 describing different paths for particular spatial positions of the source in the median plane. The reflections can be considered as delayed-and-filtered copies of the original signal. Resulting constructive and destructive interferences between the direct and reflected sound causes spectral comb filter features. Therefore, for particular positions of the sound source in the median plane, a sound with a specific combination of spectral peaks and notches reaches the entrance of the ear channel [29], [30].

Besides shoulder reflection, a similar phenomenon is triggered by reflection on the pinna rugged structure [31]. In this case, the source position in the median plane also results in the effect of comb filtering and consequent unique combination of spectral features [32]. The effect of pinna reflection is demonstrated in Fig. 2.4.

Due to different size of shoulder and pinna mass, and also due to unique delay times, specific frequency bands are affected. In relation to the wavelength of the incoming sound, shoulder reflection affects mid-frequencies and pinna reflection affects high frequencies [33]. Spatial orientation also depends on the character of the environment, where the listener is located. The above-mentioned reflections from the pinna structure and shoulder reflections and the reflections from elements of the surrounding environment specify the information about the sound source position [9], [34]. The final signal received at the position of the ear-canal entrance is a composition of the direct sound and several environment reflections which deliver specific filtering for a particular spatial configuration [35], [36]. It is worth noting that the environmental reflections interact with body structure as well. Interaction with the environment is demonstrated in Fig. 2.16. [1]

[1] For the purposes of this thesis, a convention for color description of the left and right channels is instituted; *Red* for the Right channel and *Black* for the left channel.

2.2 Basics of Virtual Sound Source Positioning

The objective of the virtual positioning process is to create an illusion of a particular position of the virtual sound source for the listener [37], [38]. Generally, the essence is to affect the information the for brain coding process in a way it will evaluate the location of the presented sound in a desired position [39]. The most intuitive approach is a natural direct reproduction of the signal from the desired position around the listener, which consequently refers to systems with multiple loudspeakers (loudspeaker arrays) [40]. These systems are widely introduced in cinemas, home cinemas, or PC gaming systems and their spatial impression[2] and their effectiveness is very robust. However, these systems have several disadvantages [41], [42]:

- fixed position of the presented multimedia scene.
- no option for separation of listeners and non-listeners.
- challenging spatial requirements.
- challenging hardware requirements (involving system costs).

Not every application using virtually positioned sound is able to fulfill all the requirements, therefore the multi-loudspeaker system solution is not acceptable. Another solution is to transform the listening situation to headphone-based listening [7], [43], as demonstrated in Fig. 2.5.

In the Figure, the listening conditions in panel (b) are limited to only two channels with *dichotic* (mutually separated) signals. Headphone-based listening is independent on specific loudspeaker arrangement (a) and can be introduced wherever by appropriate positioning process of the presented signals.

The main difference in loudspeaker-based and headphone-based systems is illustrated in Fig. 2.6. In listening conditions with real loudspeakers (demonstrated by stereo configuration) *channel crosstalk* occurs, when the signal from left loudspeaker h_L is received as in the left (h_{LL}), as in the right (h_{LR}) channel and vice versa. This phenomenon prevents the perception of the acoustic image in a wider range than what is determined by the loudspeaker distance [40]. In contrast, channel separation in headphone-based listening may significantly extend the acoustic scene, as described in the next text.

This thesis focuses on virtual positioning for headphone-based listening with no particular aim at the issue of physical multi-loudspeaker systems. Implementation of the headphone-based methods requires simulation of essential phenomena regarding localization cues, briefly described in 2.2. The following subsections describe the options on how to achieve the spatial illusion of a particular sound source.

[2] For the purposes of this thesis a phrase *spatial impression* is used for assessing fidelity of the virtual space, spatial depth, and immersion of the presented acoustic scene. More details available in Sec. 2.6

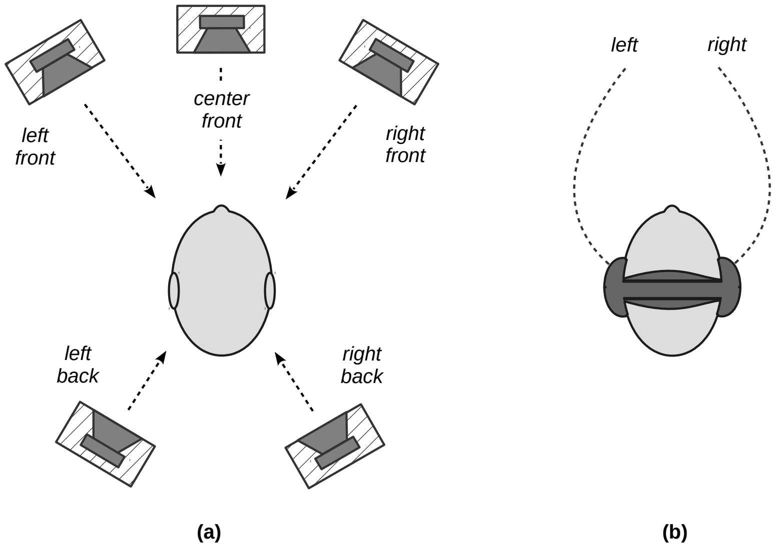

Figure 2.5: Listening systems. System with multiple real sources is depicted in panel *(a)* and headphone-based listening *(b)*.

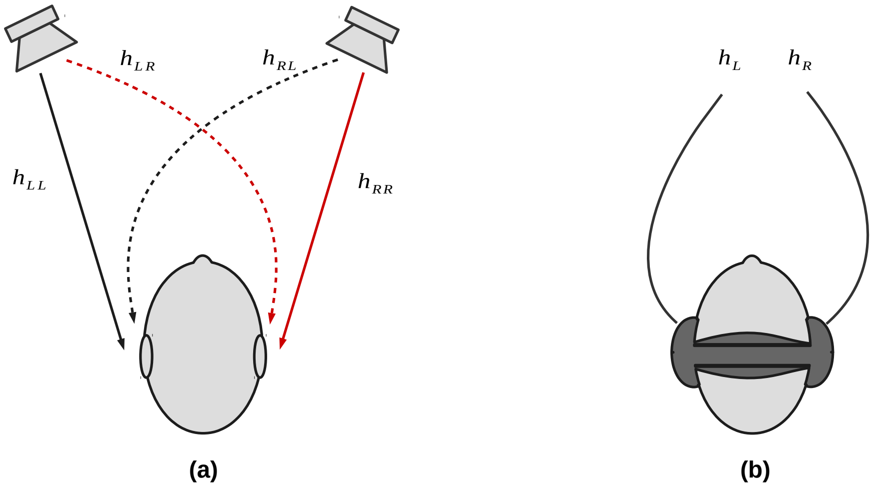

Figure 2.6: Crosstalk. The difference between listening in a real sound field with multiple sources and headphone-based listening.

2.2.1 Amplitude Panning

A common and mostly used method for creating spatial separation of the sources is known as *amplitude panning* (AP). This method has been implemented over majority of applications due to its simplicity and *sufficient* spatial impression (e.g. music industry, PC games, film production, etc.). Furthermore, this method works well both with headphone-based listening and loudspeaker-based listening. The implementation puts to relation position of the source in the horizontal plane and corresponding gain of left and right channel, therefore the ILD occurs. A widely used approximation of channel gain dependence on azimuth is known as *sine law* (SL) that offers analytical description for both channels [44]. There are several other models (e.g. *tangential law* [1], or method for bias reduction); however, for the purpose of this work, the SL primarily represents positioning approach of frequency independent gain modifications, thus a comprehensive description of all variants is considered beyond the scope. As said, *sine law* puts into relationship the position of the source in the horizontal plane and the corresponding gains of the left and right channels, described as

$$\sin(\varphi) = \frac{g_R(\varphi) - g_L(\varphi)}{g_R(\varphi) + g_L(\varphi)}, \tag{2.3}$$

where g stands for gain of corresponding channel specified by indices.[3] This equation is supported by law of constant summation of particular channel energy

$$g_L^2(\varphi) + g_R^2(\varphi) = 1. \tag{2.4}$$

The amplitude panning in its essence defines artificial ILD between both left and right channels by linear-scaled gain. In practical use, it offers sufficient ratio of complexity of implementation and spatial effect. However, this positioning method does not correspond to the real conditions at all due to the following reasons:

- There is no real frequency dependence of the ILD
- The ITD is neglected
- Limitation in frequency range of use (Sec. 2.1)
- This method allows only lateralization effect, not externalization [45]
- Spatial positioning is allowed only within the horizontal plane

As mentioned in Section 2.1, the ILD cues for sound source localization are frequency-dependent and limited for lower frequencies. This method is also unable to deliver any localization cues for the median plane. Further description of this method is available within Chapter 4, where the positioning methods are examined in real use.

[3] This work deals with the description of double channel processing which requires numerous indexing of the *left* and *right* channels. For equations where preserving the detailed channel description is not necessary (e.g. description to left channel is analogical to right) indexing by ξ is used.

2.2.2 Time Shift

Similar spatial effect, as by the ILD, is possible to reach not by adjusting the amplitude ratio of the channels, however, by adjusting the time shift between the channel signals. This approach is not used as widely as amplitude panning due to more complex implementation. Using the ITD positioning requires a model for delivering appropriate ITD values for corresponding source positions. Simple model based on approximation of the head by a sphere (circle) can be found in [46] and [28]. Demonstration of the geometry of this approximation is shown in Fig. 2.7.

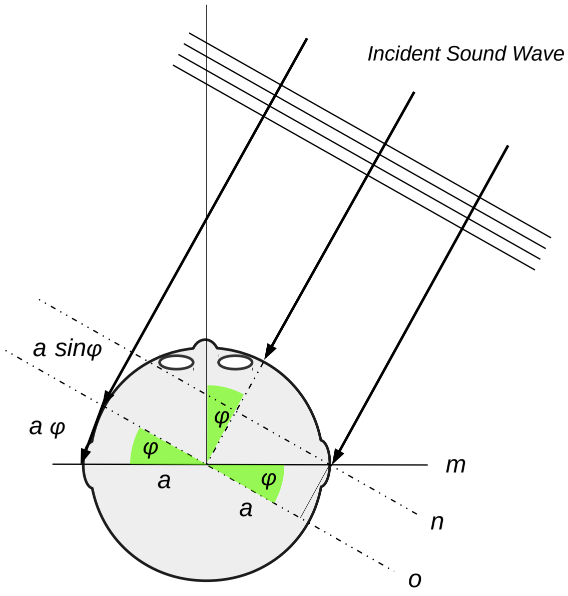

Figure 2.7: ITD geometry. Demonstration of the geometry arrangement for the ITD acquisition, when head is approximated by a sphere [46]. The extra-propagated path of the left channel consists of the direct ($a_h \sin \varphi$) and bended path ($a_h \varphi$).

Parameter a_h defines the head radius and resulting ITD range. Sound propagating from a source in azimuth φ reaches the right ear first. Line n defines equidistant lengths of the sound wave. The ITD is determined by time necessary to reach left ear. This time can be computed as a ratio of path length and speed of sound. The path consists of two parts; distance of lines n and o expressed as $a_h sin(\varphi)$, and a part of perimeter expressed as $a_h \varphi$. Therefore, the final expression of the ITD from this model referred as $\mathrm{ITD}^m(\varphi)$ can be written as

$$\mathrm{ITD}^m(\varphi) = |t_L(\varphi) - t_R(\varphi)| = \frac{a_h}{c_0}(\sin(\varphi) + (\varphi)) \tag{2.5}$$

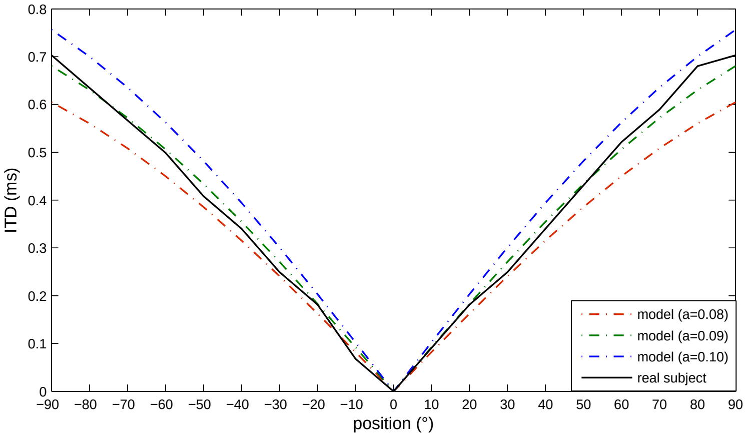

Figure 2.8: Evaluation of ITD approximation. Comparison of a real ITD (author's) with a sphere approximation of three different head radii is introduced here. Green line corresponding to $a = 0.09$ m fits the original curve the most.

Variables t_L and t_R refer to the time of incidence. The question of appropriate head size, i.e *anthropometric parameters*, is discussed in greater detail in the following subsection. According [17], average diameter of human head is in size about 16 cm. For comparison of the model efficiency, see Fig. 2.8. This figure introduces how setting of specific radius can fit a real measured ITD curve.[4] Setting of a different radius may result in perception offset, when different particular position of the source is perceived according to actual ITD of both channels.

Despite human head has rather ellipsoidal character [47], circle approximation delivers sufficient results. Moreover, to complex implementation of this algorithm, ITD positioning is limited by discrete values of delay step. In case of digital processing, this step is defined by reversed value of sampling frequency, as introduces Eq. (2.6).

$$\Delta_{\mathrm{ITD}} = \frac{1}{fs} \qquad \Rightarrow \qquad \Delta_{\mathrm{ITD}} = 22.68 \cdot 10^{-6}\Big|_{fs=44100} \quad (s) \tag{2.6}$$

For standard digital processing of 44.1 kHz, the Δ_{ITD} is determined as 22.68 μs. Maximal (average) ITD value of 0.7μs for side position $\varphi = 90°$ corresponds to delay of 31 samples. This results in theoretical resolution of approximately 3° in the horizontal plane. However, for lower sampling frequency the resolution will become insufficient. Furthermore, implementation of the ITD looses its spatial impression around 1.5 kHz analogically to the mentioned ILD [18].

As mentioned above, both the ILD and the ITD methods can not deliver spatial impression over the whole audible frequency range. The ILD is strongly frequency-dependent

[4] The real measured ITD data corresponds to the author's own ITD obtained through the processing of his measured Head-Related Transfer Function set (see 2.3)

(ITD is frequency-dependent as well, but not so strictly [9]) and neither of both provides the effect of externalization of the positioned sound [45], [2].

For immersive spatial sound, positioning by only the *Interaural Time Difference* or the *Interaural Level Difference* is not sufficient, since its frequency independence does not determine the source position conclusively for the listener. For better results, it is necessary to introduce frequency dependence and also ensure the influence of the ITD and the ILD at the same time. More sophisticated approach can be achieved by so-called *Head-Related Transfer Function*, described in the following text.

2.3 Introduction to Head-Related Transfer Function

For higher efficiency in virtual sound source positioning, it is necessary to combine the effect of the ITD and the ILD and also introduce frequency dependence of both. Generally, the more information in correspondence with real listening conditions is delivered, the higher spatial impression can be evoked. As mentioned in the previous sections, the character of sound wave body interaction in real conditions is unique over azimuth and elevation of the sound source. For particular source positions, the signal received in left and right ear contains specific delay and alteration of spectral envelope. This approach allows to consider that the signal modifications caused by the interaction are essentially a product of specific *filtering*. Therefore, a pair of unique filters belongs to each particular position in space; for left and right ear. Since the system source-listener can be considered as Linear Time Invariant (LTI) system [48], the character of filtering (i.e. sound modification) can be described by impulse response $h(t)$. An arrangement of source-listener system is described in Fig. 2.9. When the sound source emits the signal $x_m(t)$, the signal is affected during

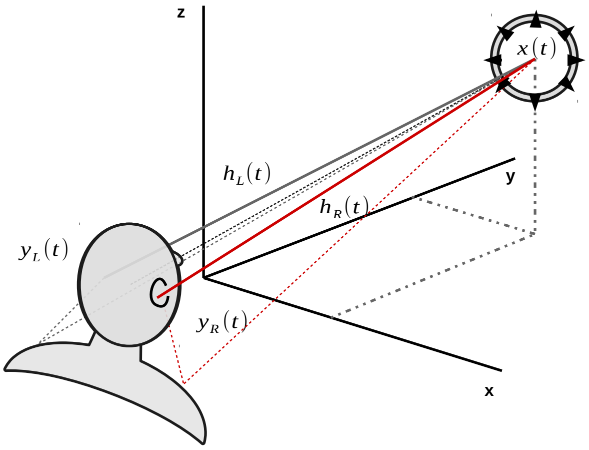

Figure 2.9: Source-listener system. The source-listener spatial configuration is considered as an LTI system. The original signal $x_m(t)$ is affected by a channel impulse response $h_\xi(t)$ resulting into a perceived signal $y_\xi(t)$.

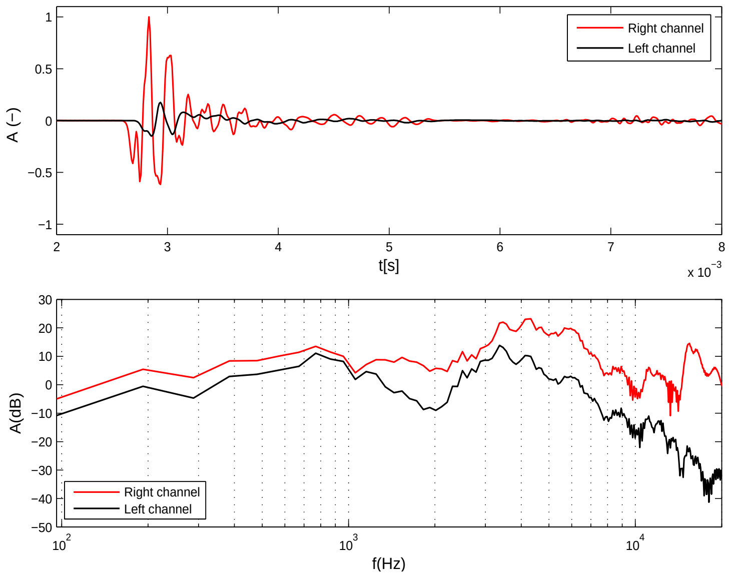

Figure 2.10: HRIR and HRTF. A pair of the HRIRs (top) for the left and right ear channels in the time domain and corresponding HRTF pair in the frequency domain (bottom).

its propagation by the transfer path described by impulse response $h_\xi(t)$, resulting into output $y_\xi(t)$ (note that ξ refers to *left* or *right* channel index, as defined above). Therefore, this *output* signal from the system is in fact the *input* signal for hearing system; i.e. signal entering the listener's ear canal entrance. The impulse response $h_\xi(t)$ is named *Head-Related Impulse Response* (HRIR) and provides the information about influence of the physical path to the propagated signal. Equivalent for the HRIR in the frequency domain is the *Head-Related Transfer Function* (HRTF) [49], [9], [50]. Both the HRIR and HRTF are basically a function of azimuth and elevation angles. Relation between the HRIR and HRTF is defined by Eq. (2.7). The transition from the time domain to the frequency domain is defined by *Fourier transform* [5]

$$\mathrm{HRTF}_\xi^{\varphi,\vartheta}(\omega) = \mathrm{FT}\{hrir_\xi^{\varphi,\vartheta}(t)\} = \int\limits_{-\infty}^{+\infty} hrir_\xi^{\varphi,\vartheta}(t)\cdot \mathrm{e}^{-j\omega t}\mathrm{d}t \qquad \xi \in (L,R), \tag{2.7}$$

where ω stands for angular frequency, FT refers to *Fourier transform*. The HRTF can be defined as a ratio of Fourier transforms sound pressure of the out signal of the system $p_\xi(t)$ and the initial spectral content of the source signal $p_s(t)$. The HRTF can be separated into modulus (magnitude), having the information about spectral content modifications, and

[5] In terms of neater equations, angle-dependence is labeled as a superscript. Therefore $hrir_\xi^{\varphi,\vartheta}(t) \equiv hrir_\xi(\varphi,\vartheta,t)$. Furthermore, for clarity, the time domain is referred by *lower case* (i.e. *hrir*) and the frequency domain by *upper case* (i.e. HRTF).

phase, having information about specific signal delay. This is summarized as

$$\mathrm{HRTF}_{\xi}^{\varphi,\vartheta}(\omega) = \frac{\mathrm{FT}\{p_{\xi}(t)\}}{\mathrm{FT}\{p_s(t)\}} = \left|\mathrm{HRTF}_{\xi}^{\varphi,\vartheta}(\omega)\right| \cdot \mathrm{e}^{j\psi_{\xi}(\omega)} \qquad \xi \in (L,R), \tag{2.8}$$

where ψ stands for phase of the HRTF spectrum and $|\cdot|$ for module acquisition. The signal received in particular ear and involving the modifications can be defined in the time domain by convolution of the signal emitted from the source $x_m(t)$ and appropriate HRIR[6]

$$y_{\xi}^{\varphi,\vartheta}(t) = hrir_{\xi}^{\varphi,\vartheta}(t) * x_m(t) = \int\limits_{-\infty}^{+\infty} hrir_{\xi}^{\varphi,\vartheta}(\tau) \cdot x_m(t-\tau)\mathrm{d}\tau \qquad \xi \in (L,R), \tag{2.9}$$

where τ refers to alternative time defining the time shift in the convolution process. The convolution procedure in the time domain defined by Eq. (2.9) can be also expressed in frequency domain as mutual multiplication of $x_m(t)$ and $hrir_{\xi}^{\varphi,\vartheta}(t)$ spectra, as[7]

$$Y_{\xi}^{\varphi,\vartheta}(\omega) = \mathrm{HRTF}_{\xi}^{\varphi,\vartheta}(\omega) \cdot X_m(\omega) \qquad \xi \in (L,R). \tag{2.10}$$

Practical demonstration of the above-mentioned definitions is shown in Fig. 2.10, which was obtained by author's own measurements [A.14]. The figure introduces a pair of the HRIRs (top) for left and right channel corresponding to source position $\varphi = 90°, \vartheta = 0°$ (i.e. right side position with no elevation), and equivalent pair of the HRTFs for the same position (bottom). Since the HRTF is relevant only for audible range, it is shown usually in this range (i.e. 20 Hz - 20 kHz) regardless other system settings (sampling frequency).

The time domain provides apparent information about mutual time shift of both channels. Onset of left channel of the HRIR is delayed by time corresponding to the ITD and energy of the response is obviously lower, resulting from the head attenuation. Lower gain of the left channel is apparent also in the frequency domain. Notice different frequency-dependent character for the left and the right ear, especially in a band between 1-2 kHz and above 15 kHz in this example. It is obvious that the higher band is affected much more than the lower band due to its inability to bend around the head. For practical applications, a whole set of the HRTF is required. Positioning by the HRTF is described in the following section.

[6] The physical character of the signal (i.e. sound pressure in (Pa)) is replaced by $x_m(t)$ as the final further description standard.

[7] Assuming that each signal in the time domain has its equivalent in the frequency domain, defined by the Fourier transform. Therefore $x_m(t) \xrightarrow{\mathrm{FT}} X_m(\omega)$.

2.4 Virtual Positioning by the HRTF

In Sections 2.2.1 and 2.2.2, the signal in each channel was affected by sample-by-sample multiplication by specific gain, or delayed by appropriate time period (number of samples). Regarding the HRTF positioning, the signal has to be processed by filter of specific characteristics. The HRTF positioning is an advanced well-known method commonly used in headphone based applications, where high-fidelity reproduction is required, e.g. virtual reality, simulators, advanced gaming [51], [52], [53], [54]. For the purposes of introduction into the HRTF theory, a continuous time was used for description. However, this thesis deals mainly with discrete signal processing, thus let the continuous time (t) to be substituted by the discrete time $[n]$. The background theory of signal processing can be found in e.g. [55], [1], [56]. Therefore, $dhrir_{\xi}^{\varphi,\vartheta}[n]$ is represented by sequence of samples of length N. As mentioned above, the HRTF method is limited only for binaural reproduction[8] through headphones, since separated channels are required. The pair of the HRTFs carries all the necessary information about signal transfer modifications, i.e. magnitude adjustment representing the ILD and phase shift representing the ITD. After delivering the same relative time and spectral features to the signal to be positioned, the hearing system evaluates its position according to these features and spatial impression occurs [9]. In terms of virtual positioning, two main approaches are used: *Static virtual positioning* and *Dynamic virtual positioning*. Both are described in the following sub-sections.

2.4.1 Static Virtual Positioning

As stated above, the HRTF can be considered as a pair of direction-dependent filters. Static virtual positioning is a specific processing, when the final virtual position of the positioned sound corresponds to one fixed position. This fact reduces its practical implementation to standard filtering of the signal by FIR filter described by the HRIR coefficients [55]. A scheme of such filter is shown in Fig. 2.11. The coefficients $h[n]$ directly corresponds

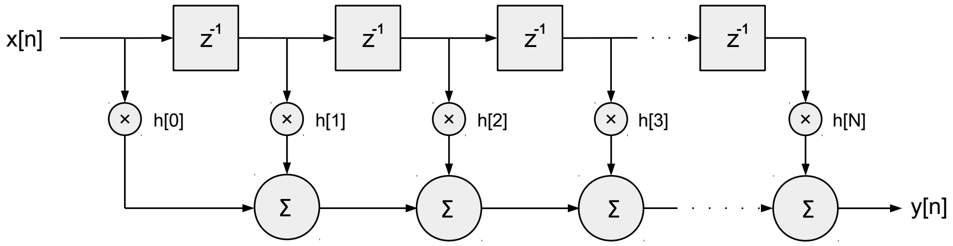

Figure 2.11: Static FIR filter structure. Structure of the filter for *static positioning* contains constant coefficients.

[8] Unaccounted approach of limiting channel crosstalk in stereo reproduction, e.g. in [40]. However, this approach is not widespread.

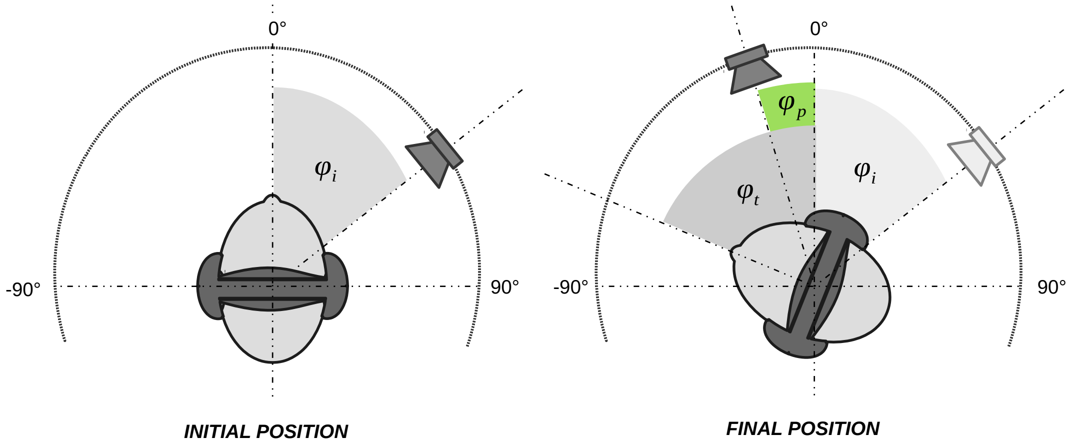

Figure 2.12: Static positioning. A scheme of its effect on the virtual sound source position with the subject's movement included is shown here. The source position is shifting along the head angle.

to samples of $hrir[n]$. The process of filtering can be expressed by transformation of Eq. (2.9) into discrete time domain. Therefore, the integral turns into summation and infinite boundaries are substituted by actual size of the signal. This process is described by Eq. (2.11) that are expressing the linear convolution[9].

$$y_{\xi}^{\varphi,\vartheta}[n] = hrir_{\xi}^{\varphi,\vartheta}[n] * x_m[n] = \sum_{k=0}^{M-1} hrir_{\xi}^{\varphi,\vartheta}[k] \cdot x_m[n-k] \qquad \xi \in (L,R), \tag{2.11}$$

where $[k]$ stands for discrete time shift and M for length of the response. Static virtual positioning creates an illusion of sound source in space with *absolute* position to the subject's head, but *relative position* to the virtual space. This effect is illustrated in Fig. 2.12. For simplicity it is reduced only to horizontal plane and the coordinates are related to absolute position of the scene, not to the listener's head.

The source is placed in initial position of virtual position φ_i. After a head turn by angle φ_t counter-clockwise, the source is perceived in a new angle φ_p despite its absolute position in the scene remains in φ_i. Therefore, subject's head movement changes the source position to a new one, where $\varphi_p = \varphi_i + \varphi_t$. Generally, the subject instantly carries the virtual scene embedded in his head. It depends on particular application, whether this effect is desirable or not [53], [57], [58]. Static positioning can be required for specific applications, e.g.:

- Static virtual reality, e.g. sound track for specific visual presentation, where acoustic track does not formally follow the visual content.

[9] Linear and circular convolution for the implementation of the static and dynamic sound source are fully described in Section 4.2 and in [A.4].

- Virtual separation of the sound source, e.g. spatial mixing of musical record, where spatial division of the whole presented scene is more important as altogether than its absolute position.

- Experiments within real and virtual static sound source localization, where directly this features are about to be explored.

For other applications, rather dynamic virtual positioning is necessary to be implemented.

2.4.2 Dynamic Virtual Positioning

For several applications, a fixed position of the source is insufficient. Therefore, such applications require implementation of moving sound source. Dynamic virtual positioning is a specific processing, where the final virtual position of the positioned sound varies. This effects is desirable in three main cases:

1. The scene fixed to listener's head with *relative* spatial position of the source requires moving source (e.g. spatial music effect, or movie sound track).

2. The scene with *absolute* position of the source includes source with fixed position. The system compensates listener's head movements (body movements possible as well), therefore, according to maintaining the absolute position, the position is changed (e.g. virtual reality with cave-like settings [59], [60]). This effect is demonstrated in Fig. 2.13.

3. Combination of the above-mentioned options, when the *absolute* position in space is varying, and also head movements are compensated.

Dynamic positioning is often implemented in systems with *head-tracking* sensor [57], [43], [A.12]. This system allows to capture head movements in order to correct the source position. Therefore, in correspondence with real listening situation, after head turn of the listener, the source position is recalculated, thus *absolute* position of the source in space is maintained. The effect of *absolute* source position is demonstrated in Fig. 2.13. Note also static positioning in Fig. 2.12.

In this example, the initial position of perceived source position φ_i remains the same in absolute position in the coordinates after a head turn by angle φ_t. However, in relation to the listener, the source is now perceived in a new position φ_p. The implementation of dynamic virtual positioning is not as easy-to-implement as static positioning. The dynamic positioning can be expressed by FIR filter with varying coefficients, as shown in Fig. 2.14. The coefficients $h[n]$ directly corresponds to samples of $hrir[n]$ as in previous example. However, all to coefficients are a function of discrete time τ that introduces changes in position of the source.

The implementation of the filter shown in Fig. 2.14 is a complex approach and can be reached by a combination of several techniques. The background of implementation

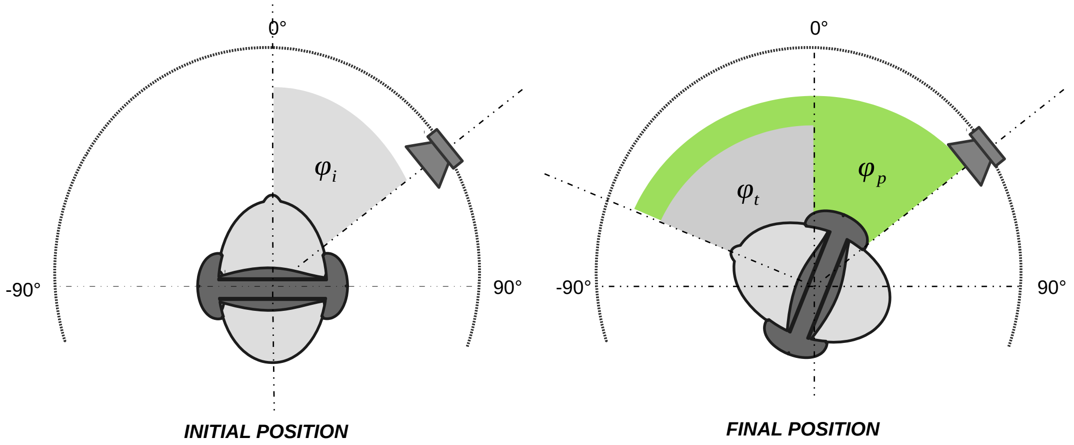

Figure 2.13: Dynamic positioning. A scheme of the head movement compensation and its impact on virtual localization with subject's head movement is depicted. The source position remains fixed.

exceeds a range of this chapter. Since the dynamic positioning algorithms are significant part of this thesis, the issue is described in more details in Section 4.2.

In summary within static and dynamic positioning, see Fig. 2.15. This scheme describes the main idea of double-channel processing issue of positioning by the HRTF. The original discrete representation of monaural signal to be positioned $x_m[n]$ is doubled and each channel processed by $hrir_{\xi}^{\varphi_k,\vartheta_k}[n]$. The appropriate impulse response is selected according to source position (which may in general vary or not). Both processed channels are finally combined to produce a positioned stereo file.

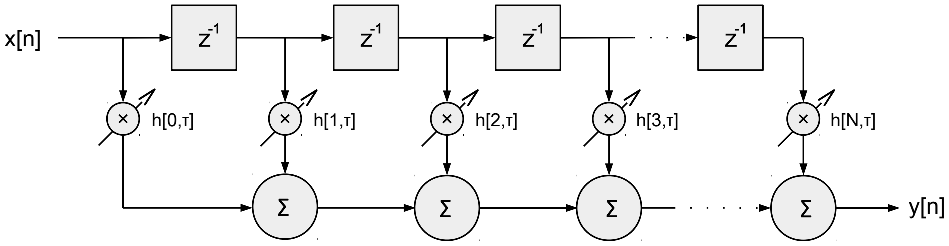

Figure 2.14: Dynamic FIR filter structure. This scheme shows the filter for *dynamic positioning* with varying coefficients $h[n]$ as a function of discrete time τ.

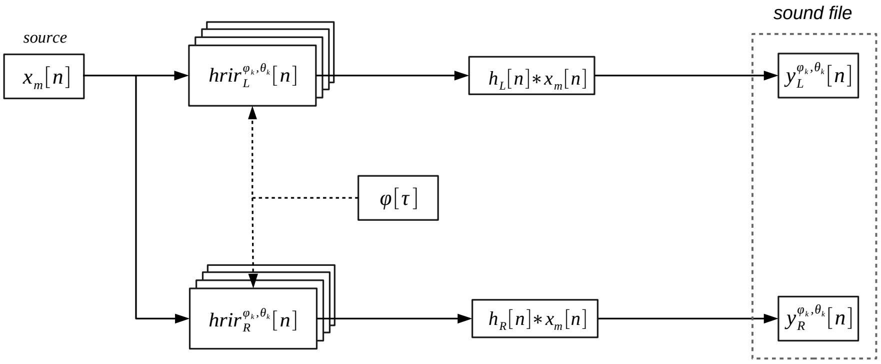
sound file
source
$x_m[n]$
$hrir_L^{\varphi_k,\theta_k}[n]$
$h_L[n]*x_m[n]$
$y_L^{\varphi_k,\theta_k}[n]$
$\varphi[\tau]$
$hrir_R^{\varphi_k,\theta_k}[n]$
$h_R[n]*x_m[n]$
$y_R^{\varphi_k,\theta_k}[n]$

Figure 2.15: HRTF processing scheme. A conceptual scheme of data processing in virtual sound source positioning by Head-Related Transfer Function is shown.

2.4.3 Environmental Reverberation

The previous algorithms assume direct propagation of the sound form the source towards the listener with no additional reflections by surrounding environment involved. For several applications, it is advisable to emulate particular acoustic character of specific environment (e.g. wall reflections, reverberant spaces, etc.). According to [35] and [36], this process can also enhance the spatial impression and extend perceived space depth. For demonstration see again Fig. 2.16. The wall reflection can be considered as the second artificial source situated in a different position than the original one. In terms of static positioning, this effect can be described by Eq. (2.12)

$$y_\xi^{\varphi,\vartheta}[n] = \overbrace{x_m[n]*hrir_\xi^{\varphi_1,\vartheta_1}[n]}^{\text{Direct signal}} + \overbrace{x_m[n-d]*h_r[n]*hrir_\xi^{\varphi_2,\vartheta_2}[n]}^{\text{Environment reflection}}, \tag{2.12}$$

where $hrir_\xi^{\varphi_1,\vartheta_1}[n]$ corresponds to the original direction and $hrir_\xi^{\varphi_2,\vartheta_2}[n]$ represents the direction of wall reflection. Constant d defines the delay of the reflected signal and $h_r[n]$ represents frequency-dependent attenuation of the wall. However, in the real environment multiple reflections occurs. For spatial orientation, the first several reflections with higher energy are important [2], [9]. Multiple environment reflections can be involved by modification of Eq. (2.12) the following equation

$$y_\xi^{\varphi,\vartheta}[n] = \overbrace{x_m[n]*hrir_\xi^{\varphi_1,\vartheta_1}[n]}^{\text{Direct signal}} + \overbrace{\sum_{i=2}^{M} x_m[n-d_i]*h_{r,i}[n]*hrir_\xi^{\varphi_i,\vartheta_i}[n]}^{\text{Multiple environment reflections}}. \tag{2.13}$$

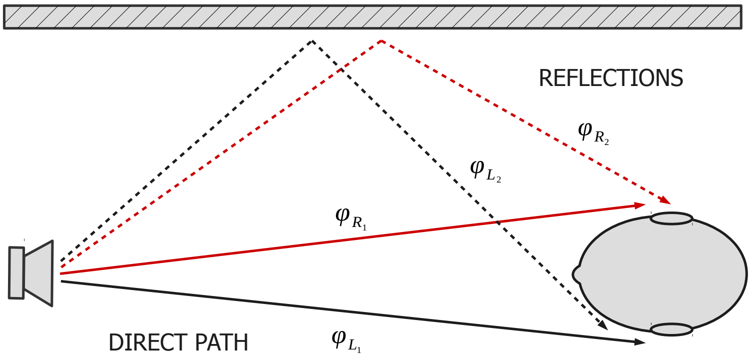

Figure 2.16: Environmental reverberance. Sound propagation in a real environment results with reflections is demonstrated. The environment reflections are propagating from distinct directions.

2.5 Attributes of HRTF set and HRTF acquisition

The data regarding position-dependent modifications has been delivered by *Head-Related Transfer Function*. However, there exists several options of obtaining an appropriate HRTF set. Series of tasks have to be challenged, the highlights are mentioned below:

- For universal use of the HRTF, it is necessary to sample particular spatial positions for the whole relevant perimeter surrounding the listener. Therefore, the HRTF is mostly delivered for discrete positions with *sufficient* density. In real use, more over 500 pairs of the HRTF is needed [17], [61].
- The HRTF is strongly dependent on *anthropometric* parameters of the subject (i.e. head diameter, head height, torso width, pinna surface, etc.) [62], [63]. Therefore, each listener should use his own set of the HRTF under ideal conditions [64], [65]. In case of using set of someone else, a disturbing offset in localization and confusion of the subject may occur [66].

The most usual way of the HRTF acquisition is either direct measuring on a subject [67] [68], or modeling according to the anthropometric parameters [69], [70]. The modeling can also involve various approaches. Both attitudes are discussed in the following sections.

2.5.1 Measuring of HRTF

Resulting from *signal-and-systems* theory, impulse response (i.e. transfer function) can be introduced for LTI systems: the impulse system of unknown LTI system can be obtained by appropriate measuring method based on input-output relation, when the specific output

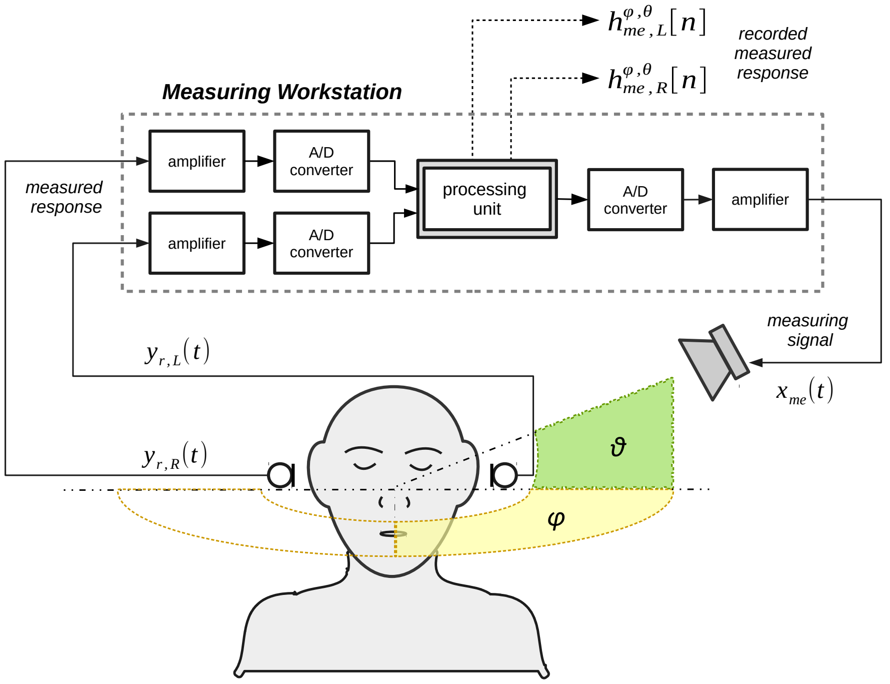

Figure 2.17: HRTF measurement settings. The output discrete signal $h^{\varphi,\vartheta}_{me,\xi}[n]$ corresponds to a recorded response of the source-listener system $y_{r,\xi}(t)$ to the measuring signal $x_{me}(t)$.

of the system is compared to the input [48]. The similar situation is also the HRTF measurement. Measuring of the HRTF involves a use of two miniature microphones attached (or likewise delivered) to both subject's ear canal entrances to record the output.[10] Since the body structure is actually the system itself with its inputs in the ear canal entrances, only a source of the measuring signal for various positions is needed. A brief scheme of simple measuring set is depicted in Fig. 2.17.

In practical measurement settings, the subject is sitting on positionable chair or surrounded by positionable loudspeaker(s) [17], [61], [68]. The two microphones send the measured signal to a measuring software, usually with use of pre-amplifiers. The measuring loudspeaker is also connected to the measuring software [A.18], [A.7]. For particular position φ, ϑ the measuring signal $x_{me}t$ (e.g. MLS, sweep-sine, etc.) is emitted from the loudspeaker and recorded as $y_{r,\xi}$ by the measuring microphones inserted in subject's ears [71]. [11] The measured distance is often normalized to 1 meter form the imaginary head

[10] This statement might appear confusing, since the *input* signal entering each ear channel is essentially the *output* signal of the system of the HRTF describing the interaction between sound and body structure. Notice the difference.

[11] The position of the microphone may vary according to the particular definitions of the HRTF [72]. The most common position of the microphones is directly at the entrance of the ear canal. This is suitable to avoid a double-canal effect.

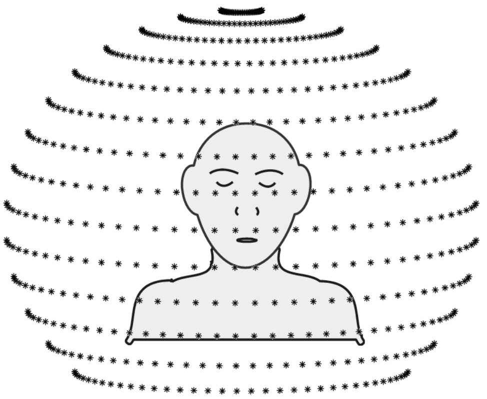

Figure 2.18: Example of HRTF measuring points. The points are often uniformly distributed with measuring step of approximately 5°.

center. After processing of the input and output signal a pair of discrete system responses $h_{me,\xi}^{\varphi,\vartheta}[n]$ is obtained. More detailed information about measuring methods can be found in the references.

The efficiency of the measurement significantly depends on the available equipment, since the HRTF for multiple positions is supposed to be measured. This may appear quiet time-consuming. Professional measuring systems allow to reach very precise density; however, the more dense the measurement is, the more complex is consequent data handling. In [17] or [68], a special measuring device is introduced. This system consists of a ring with multiple speakers on the edge, which allows multiple measurements while only one spatial configuration of the whole system is set. However, the measuring itself is very time-consuming even with appropriate equipment (in range of tens of minutes). In such systems the range between neighboring positions is usually chosen from 3 to 8 [61] and often depends also on the region, since *Just Noticeable Difference* in localization (i.e. the most precise spatial resolution) is much greater in the area of interest (i.e. directly in front of the user) than on the side positions [13], [9]. A scheme of possible measured positions in real use is demonstrated in Fig. 2.18.

Since the physical measuring equipment is not ideal, the measured HRTF is more or less affected by particular elements of the measuring chain [A.7]. Assuming each element introduces its transfer function to the final signal, the measured transfer function can be expressed as in Eq. (2.14)

$$H_{me,\xi}^{\varphi,\vartheta}[\Omega] = \prod_{j=1}^{J} H_{j,(\xi)}^{\varphi,\vartheta}[\Omega], \tag{2.14}$$

where $H_{j,\xi}[\Omega]$ represents the j^{th} element of the measuring chain including the original HRTF sequence. The Ω stands for discrete frequency domain.[12] Generally, the transfer

[12] Assuming that each signal in the discrete time domain has its equivalent in the discrete frequency domain defined by the Fourier transform. Therefore $x_m[n] \xrightarrow{DFT} X_m[\Omega]$.

functions are considered position-dependent, as results from the indices. The elements of measuring chain with the most significant impact are:

- Transfer function of the measuring loudspeaker
- Transfer function of the measuring microphone
- Additional reflections of a room, where the measurement is performed

Assuming these three elements as the most significant and also that the only significant position-dependence occurs within reflections of the environment, Eq. (2.14) can be specified and modified into time domain as expresses Eq. (2.15)

$$h_{me,\xi}^{\varphi,\vartheta}[n] = hrir_{\xi}^{\varphi,\vartheta}[n]{*}h_{r,\xi}^{\varphi,\vartheta}[n]{*}h_l[n]{*}h_m[n], \tag{2.15}$$

where $h_{me,\xi}^{\varphi,\vartheta}[n]$ refers to the *raw* measured HRIR, $hrir_{\xi}^{\varphi,\vartheta}[n]$ stands for the original non-distorted HRIR, $h_{r,\xi}^{\varphi,\vartheta}[n]$ for position-dependent room response, $h_l[n]$ for loudspeaker transfer function, and $h_m[n]$ for microphone transfer function. Equalization of the HRTF (i.e. reducing the influence of other elements) can be realized by deconvolution in spectral domain as states in Eq. (2.16)

$$hrir_{\xi}^{\varphi,\vartheta}[n] = \mathrm{FT}^{-1}\left\{\frac{H_{me,\xi}^{\varphi,\vartheta}[\Omega]}{H_{r,\xi}^{\varphi,\vartheta}[\Omega]\cdot H_l[\Omega]\cdot H_m[\Omega]}\right\}, \tag{2.16}$$

where all the transfer functions corresponds to an equivalent impulse responses in Eq. (2.15) defined by appropriate indices. Another option how to reduce the influence of room is limitation of the measured HRIR. As mentioned in previous sections, the HRIR contains several reflections of the incident sound from the body structure. The time of incidence of the last reflection resulting from body interaction is supposed to be shorter than the time of incidence of the first reflection resulting from room interaction. Therefore, the additional reflections can be removed by limitation of the HRIR length. The number of samples of the HRIR is defined by the time needed to traverse the shortest path of the first environment reflection s_{min}, as expressed in Eq. (2.17)

$$L_{hrir} < \frac{s_{min}}{c_0}\cdot f_s \quad . \tag{2.17}$$

Reduction of room response can be expressed as selection of particular range of the digital signal, therefore[13]

$$hrir_{\xi,lim}^{\varphi,\vartheta}[n] = \Big(hrir_{\xi,me}^{\varphi,\vartheta}[n]\Big)_{0:L_{hrir}} . \tag{2.18}$$

The whole process of rendering virtually positioned stereo file from monaural source including appropriate limitation an equalization of the HRTF set is demonstrated by block diagram in Fig. 2.19.

[13] In this thesis, selected segments of the digital signal are expressed as $x_2[n] = \Big(x_1[n]\Big)_{a:b}$. This stands for the selection of signal $x_1[n]$ from a^{th} sample to b^{th} sample.

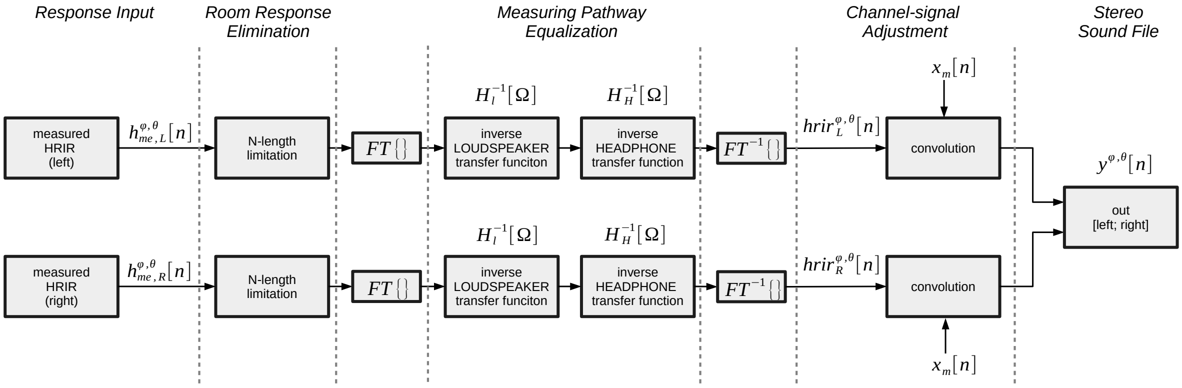
Response Input
Room Response Elimination
Measuring Pathway Equalization
Channel-signal Adjustment
Stereo Sound File
measured HRIR (left)
$h^{\varphi,\theta}_{me,L}[n]$
N-length limitation
$FT\{\}$
$H_l^{-1}[\Omega]$
inverse LOUDSPEAKER transfer funciton
$H_H^{-1}[\Omega]$
inverse HEADPHONE transfer function
$FT^{-1}\{\}$
$hrir^{\varphi,\theta}_L[n]$
$x_m[n]$
convolution
$y^{\varphi,\theta}[n]$
out [left; right]
measured HRIR (right)
$h^{\varphi,\theta}_{me,R}[n]$
N-length limitation
$FT\{\}$
$H_l^{-1}[\Omega]$
inverse LOUDSPEAKER transfer funciton
$H_H^{-1}[\Omega]$
inverse HEADPHONE transfer function
$FT^{-1}\{\}$
$hrir^{\varphi,\theta}_R[n]$
convolution
$x_m[n]$

Figure 2.19: Positioning of a stimulus. Synopsis of a virtually positioned stimuli with additional processing of the measured HRTF, including limitations in the discrete time domain and equalization in the frequency domain.

2.5.2 Modeling of HRTF

A disadvantage of direct measuring of the HRTF is that a net of measured points with sufficient density is needed. Therefore, it requires a lot of time spent with measuring process multiplied by series of subjects. Another way how to obtain the HRTF set is to implement a mathematical model based on simply measurable anthropometric parameters [14]. Models are constantly improved with good results, however, in terms of sound source perception, the quality is still worse than measured set [12], [29], [73], [70], [74]. The field of HRTF modeling is very complex, containing various approaches (structural modeling, mathematical approximations, synthesis by boundary element method - BEM, etc.). For demonstration purposes the model published in [46] is introduced. This model is based on description of multiple paths that the sound travels from source to listener's ear. A scheme for left and right ear is depicted in Fig. 2.20. Both channels of the model work on the same principles; the main interactions of the sound wave and the body structure are simulated.

Each reflection (see Fig. 2.3 and Fig. 2.4) in this model is represented by coefficient ρ, which determines the ratio between energy of incident and reflected sound, and by coefficient τ or T representing the time delay resulting form the reflection and related to the direct sound. When the head shadowing is activated (indirect path between source and ear for contra-lateral side), ITD is set between both channels. This way delivers specific combinations of τ and ρ coefficients for all directions. Therefore, the final signal for one channel consists of sum of the direct sound and its attenuated and delayed copies. This process delivers a unique filtering of the original signal (HRTF effect) because of positive and negative frequency interferences resulting from mutual phase shift. More details can be found in e.g. [75], [28], [76], [77], [74].

Another method how to obtain an appropriate HRTF is also *scaling* or *fitting* set of already measured or modeled HRTF set. This method is based on the fact that the most

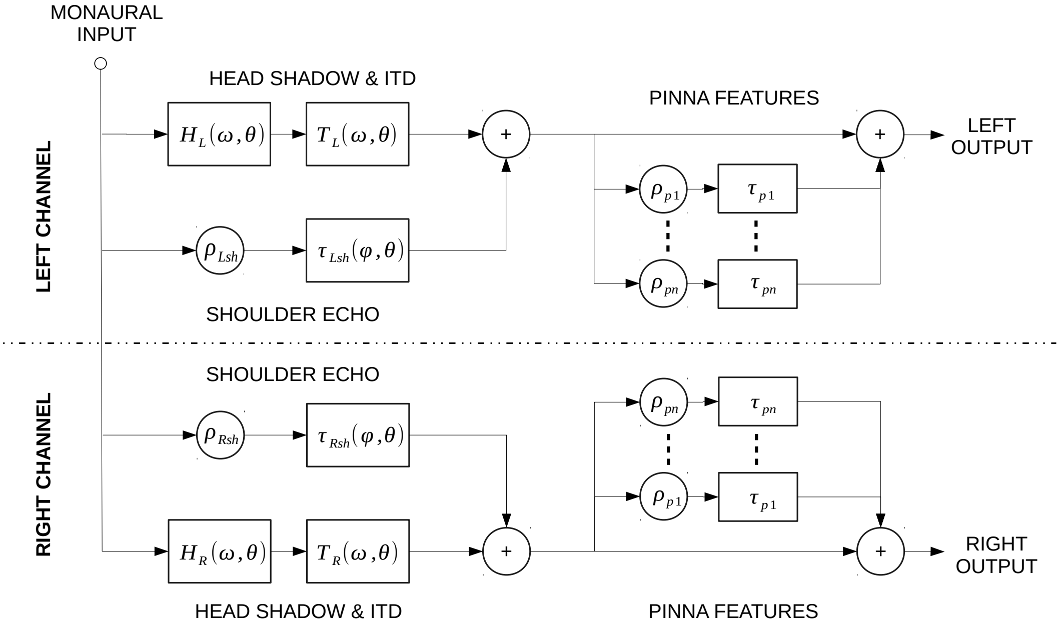

Figure 2.20: Model of the HRTF. A model of the Head-Related Transfer Function covering all the major features of sound-body interaction is demonstrated; shoulder reflection, head shadowing and pinna reflections [46].

significant spectral features of the HRTF (i.e. peaks and notches) are directly connected to the specific anthropometric parameters [78], [63], [65], [31]. Therefore, these methods either re-scale the HRTF over the frequency axis, or merging together different frequency bands from multiple sets of different subjects.

2.5.3 Representation of a HRTF set

In the previous text, the HRTF was referred to as a pair of transfer functions. For specific purposes when referred to, *HRTF set* it should be understood as a pair of matrices corresponding to appropriate ear. Each row in the matrix represents the HRIR for particular position. Therefore, this can be expressed as in Eq. (2.19)

$$\boldsymbol{hrir}^{\varphi}_{\xi,n} = \begin{pmatrix} hrir^{\varphi_1}_{\xi}[n] \\ hrir^{\varphi_2}_{\xi}[n] \\ \vdots \\ hrir^{\varphi_M}_{\xi}[n] \end{pmatrix} = \begin{pmatrix} h^{\varphi_1}_{\xi,1} & h^{\varphi_1}_{\xi,2} & \cdots & h^{\varphi_1}_{\xi,N} \\ h^{\varphi_2}_{\xi,1} & h^{\varphi_2}_{\xi,2} & \cdots & h^{\varphi_2}_{\xi,N} \\ \vdots & \vdots & \ddots & \vdots \\ h^{\varphi_M}_{\xi,1} & h^{\varphi_M}_{\xi,2} & \cdots & h^{\varphi_M}_{\xi,N} \end{pmatrix}, \tag{2.19}$$

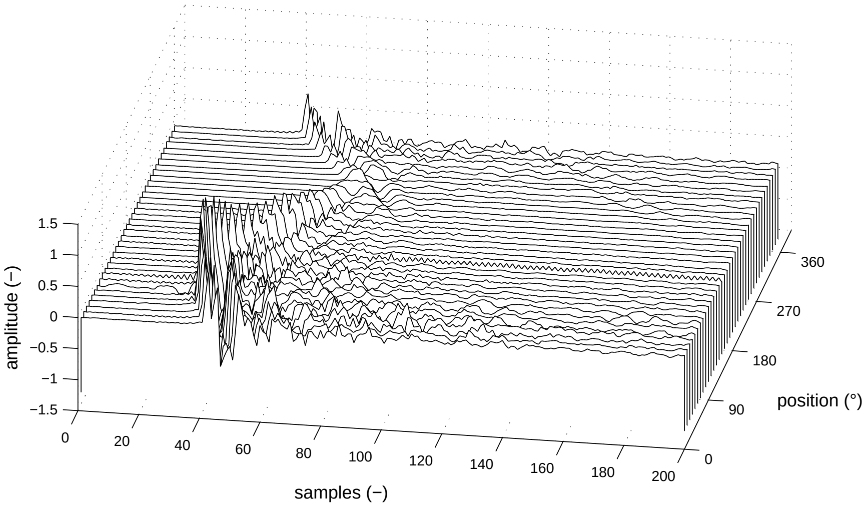

Figure 2.21: HRIR for horizontal plane. Head-Related Impulse Response of right ear for a 360-degree radius in the horizontal plane is depicted. Notice the variance of the onset time and variance of the energy of the response.

where M stands for number of positions and N for length of the HRIR, and in parallel to the time domain in frequency domain as well as expressed in Eq. (2.20)

$$\mathbf{HRTF}^{\varphi}_{\xi,n} = \begin{pmatrix} \mathrm{HRTF}^{\varphi_1}_{\xi}[\Omega] \\ \mathrm{HRTF}^{\varphi_2}_{\xi}[\Omega] \\ \vdots \\ \mathrm{HRTF}^{\varphi_M}_{\xi}[\Omega] \end{pmatrix} = \begin{pmatrix} H^{\varphi_1}_{\xi,1} & H^{\varphi_1}_{\xi,2} & \cdots & H^{\varphi_1}_{\xi,N} \\ H^{\varphi_2}_{\xi,1} & H^{\varphi_2}_{\xi,2} & \cdots & H^{\varphi_2}_{\xi,N} \\ \vdots & \vdots & \ddots & \vdots \\ H^{\varphi_M}_{\xi,1} & H^{\varphi_M}_{\xi,2} & \cdots & H^{\varphi_M}_{\xi,N} \end{pmatrix}. \tag{2.20}$$

This format of data allows to handle spatial information for one slice of the spherical coordinate system, which is useful for particular analysis.[14] Visualization of such matrix allows to observe the course of specific features representing particular localization cue. Demonstration in Fig. 2.21 shows the HRIRs corresponding to right ear for the whole range in *azimuth* plane, where $\varphi \in (0, 360)$ with step of 10°.[15]

Several specific features appear in each set of the HRIR (HRTF) regardless dependence on subject anthropometry. Onset time of the response is shortest for position $\varphi = 90°$ (source directly in front of the right ear) and also energy of the response is the highest for this position. For the following positions, as the sources shifts towards the left ear, the

[14] However, for three-dimensional positioning a use of 4-D matrices is necessary; *azimuth, elevation, number of samples, sample value.* This can be realized by e.g. several layers of 3-D matrices.

[15] The data used for demonstration in this section is author's own HRTF set measured within International Conference on Auditory Display 2013 in Lodz University of Technology, Poland

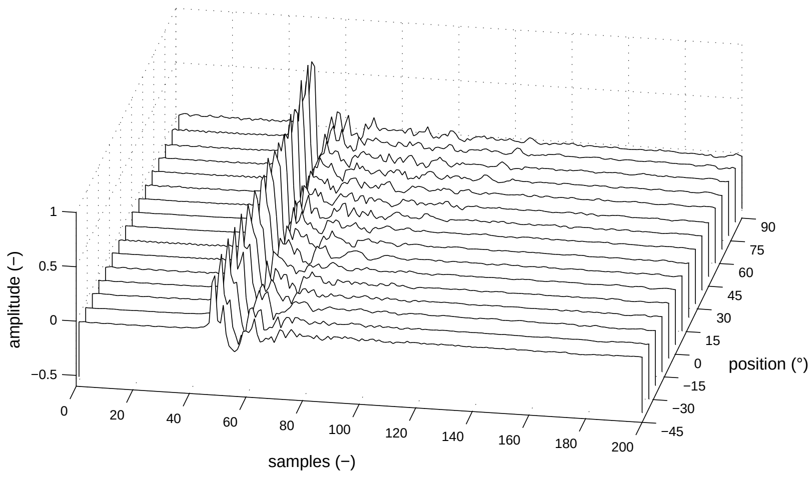

Figure 2.22: HRIR for median plane. Head-Related Impulse Response for a 135-degree radius in the median plane is depicted. Notice the same onset time and shoulder reflection wave, which is apparent for positions with a higher elevation angle.

energy of the response weakens according to head shadowing and also time of incidence extends. The effect of pinna shadowing for higher frequency bands is apparent for source positions behind the listener ($\varphi \in (90, 270)$). In comparison to the *front* positions the higher frequencies are limited resulting in more smooth character of the response. In comparison to the set for the *horizontal* plane in the previous figure, see also set for the *median* plane in Fig. 2.22. The set for median plane is limited for $\vartheta \in (-45, 90)$ since technical realization of the measuring equipment did not allow to measure the lower positions. In comparison to the *horizontal* plane the same onset time of all the responses is evident, since for the source position on this plane no ITD occur. The main difference is apparent in length of the response tail. This tail results from the shoulder reflections and is particularly apparent for $\vartheta > 30°$.

The corresponding HRTFs for horizontal and median planes are shown in representation of heat map in Fig. 2.23 and Fig. 2.24, where color represents module of the HRTF. The frequency scale is kept linear in order to emphasize frequency-dependent spectral features. Notice characteristic frequency damping for contra-lateral position in the horizontal plane (as the sound is shadowed by listener's head) and distinct spectral peaks and notches changing their position according to azimuth of the source. In median plane, larger content of higher frequencies is distinctive for increasing elevation. More comprehensive details about the HRTF features analysis can be found in e.g. [17], [79], [25], [9], [49].

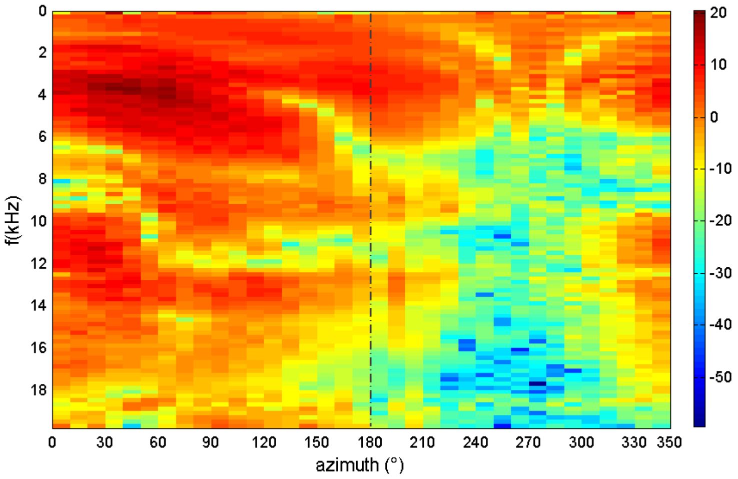

Figure 2.23: HRTF for the horizontal plane - heat map. Heat map for the Head-Related Transfer Function for a 360° radius in the horizontal plane. The color bar denotes the magnitude of the transfer function in dB.

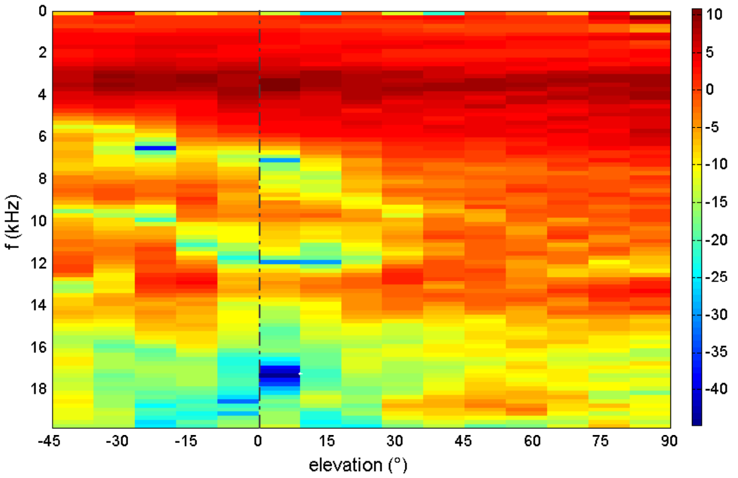

Figure 2.24: HRTF for the median plane - heat map. Heat map for the Head-Related Transfer Function for $\vartheta(-45, 90)$ in the median plane.

2.6 Quality Assessment of the Virtual Acoustic Space and its Specific Features

The previous sections summarized the principles of the HRTF (HRIR) and also introduced several methods how to obtain a specific HRTF set for creating Virtual Acoustic Space (VAS) [42], [80], [81]. The VAS can be generally considered as an audio element of virtual reality with particular virtually positioned sources. This chapter briefly introduces methods for assessing and rating the fidelity and quality of VAS (i.e. maximal correspondence with perception in real listening environment) and also specific related phenomena. However, quality of particular virtual sound source positioning is determined mostly by the final character of spatial perception of the subject regardless for instance all the objective parameters of particular HRTF sets.

A general scheme of system for rendering VAS is demonstrated in Fig. 2.25. Data of source type x_m and its position φ, ϑ is sent form a section handling the parameters of virtual scene and virtual sound sources. The head-tracking system provides correction of the source position $\Delta\varphi, \Delta\vartheta$ indicating instantaneous position of the head.[16] According to available data, the processing units delivers two signals y_L and y_R presented as a stereo file in the headphone system. One of the most relevant factors in spatial hearing allowing objective rating is *Minimum Audible Angle* (MAA), also referred as *Just Noticeable Difference*, or *Difference Limen* [13], [82], [9]. This indicator is defined as the smallest detectable difference between two spatial positions of the identical sound source. In spherical coordinate system, the JND can be defined both for horizontal plane and median plane. The principle JND is demonstrated by Fig. 2.26. Despite $\Delta\varphi_L$ and $\Delta\varphi_R$ are actually the same, the JND is generally vary across the positions [13]. It depends on the parameters of positioned signal (bandwidth, central frequency, energy, transient, etc.) as well. This can be summarized as

$$\Delta\varphi_l \approx \Delta\varphi_r = f(\varphi, \vartheta, L_{\mathrm{SPL}}, \omega, \cdots). \tag{2.21}$$

The smallest the JND is, the more finer resolution of the spatial locations is.[17] At the area of interest (in front of the user), where the localization is precise the most, the JND drops towards 1-2°,and in outer positions rises up to 10-20° [13], [9]. It is known the JND for real sources is slightly lower than for virtual sources [82].

Exploring attributes of particular applications or positioning method often requires measuring of accuracy in subject's source position evaluation rather than exact precision of JND. This fact is underlined by the main purpose of virtual auditory space; to create an illusion of spatial source location. In the application, it is important, whether the perceived position of the source corresponds to the position, determined by particular positioning method, or whether a kind of offset or confusion occurs. The metric of localization precision

[16] The sensors can also capture actual body position, which depends on the specific character of the VAS system.

[17] JND is generally defined for various factors of cognitive perception.

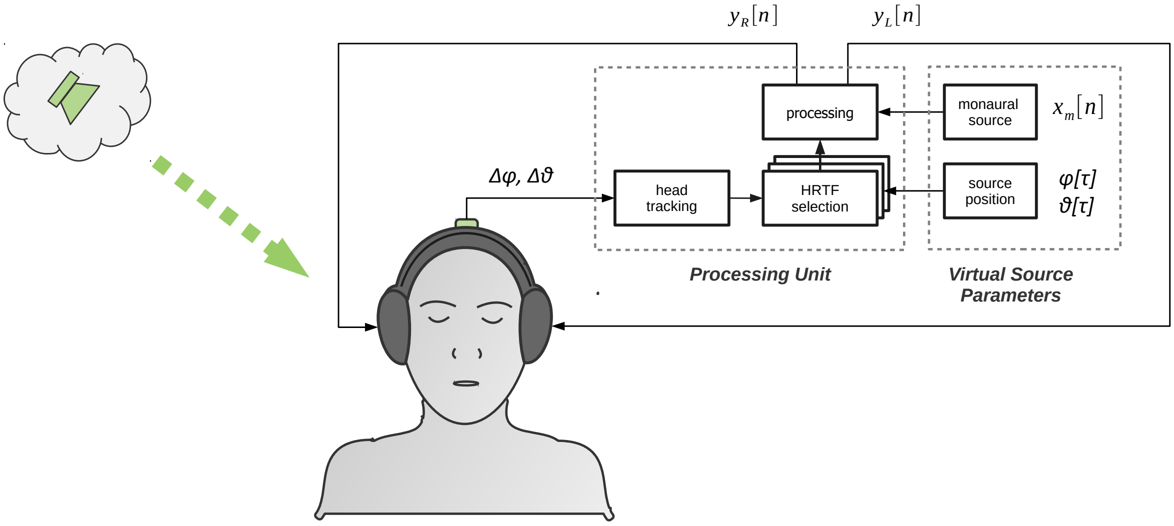

Figure 2.25: Rendering simple VAS. Scheme of data flow in headphone-based virtual positioning with head-tracking system included. Presentation of the processed signal creates an illusion of a particular sound source situated in space (*cloud*).

can be denoted as *Average Localization Error* (ALE) [A.15], expressed as follows

$$\mathrm{ALE}^{\varphi,\vartheta} = \sqrt{\frac{1}{K}\sum_{i=0}^{K}(d_i^c - d_i^p)^2}, \tag{2.22}$$

where K stands for number of test trials, d_i^c is i^{th} correct position and d_i^p is i^{th} perceived position. In general, ALE shows similar positional dependence as JND, therefore the localization error is the lowest in the area of interest. With discrete positions of uniform distribution (e.g. for purposes of subjective tests, where the subject assigns the perceived position to the offered ones) ALE can be normalized by size of the position step, therefore

$$\mathrm{ALE}_n^{\varphi,\vartheta} = \frac{1}{\Delta\gamma}\sqrt{\frac{1}{K}\sum_{i=0}^{K}(d_i^c - d_i^p)^2}, \tag{2.23}$$

where γ stands for either angle in horizontal or median plane. Normalization allows comparison of results of different settings of spatial distribution. ALE allows efficient comparison of particular positioning methods.

Besides JND and ALE, which are metric coefficients of rating in terms of VAS, several other attributes and perceptional features are useful for sufficiency of virtual positioning method.

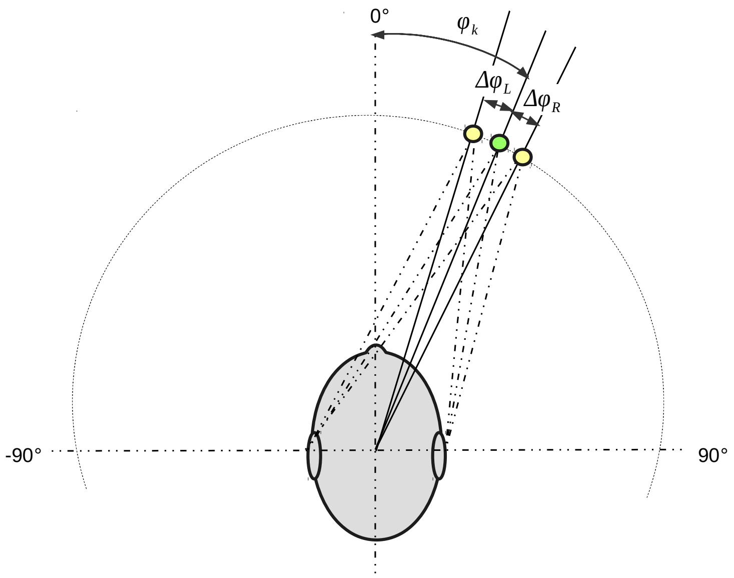

Figure 2.26: Horizontal plane JND. A scheme for the Just Noticeable Difference (JND) for the horizontal plane. For desired location of angle φ_k, the source can be shifted either to the left or right direction by $\Delta\varphi$. Generally, it is considered that $\Delta\varphi_L \approx \Delta\varphi_R$.

- **Externalization**

  Perception of externalized sound determines whether the sound source is perceived inside or outside the head. It is directly connected with *spatial impression*, which determines the width and depth of the VAS [A.2]. This effect is most important for plasticity and credibility of the illusion of the positioned source. The opposite of *externalization* is referred to as *lateralization*, when the source is perceived inside-the-head on the junction of both ears. Both variants are demonstrated in Fig. 2.27.

- **Cone of Confusion**

  Within modeling of the HRTF, certain approach involved approximation of head by a rigid sphere [28], [75]. Due to the approximation, when the artificial *ears* are placed symmetrically on the axis of the head, specific area with the same ILD and ITD emerges. This area is called *cone of confusion* [9], [3]. As demonstrated in Fig. 2.28, all positions on the surface of the cone corresponds to the same ILD and ITD. The cone is determined by one angle; here demonstrated as φ_1 and φ_2. Within a real head shape, the *cone of confusion* is generally not a regular cone. The situation is frequency-dependent and more complex due to presence of pinna and specific facial features. This has an effect on particular band widths.

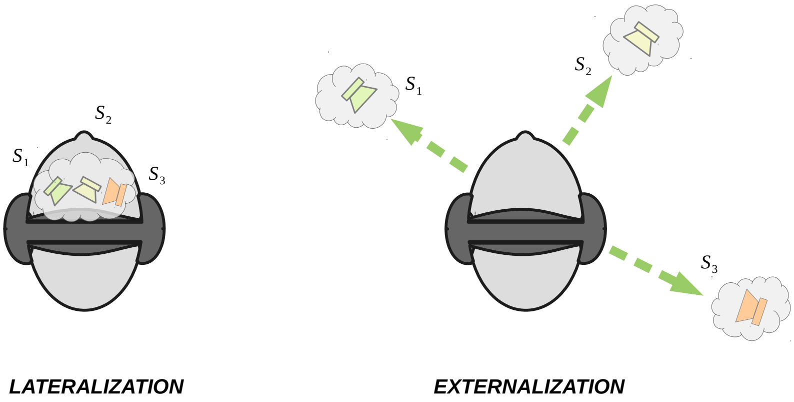

Figure 2.27: Lateralization vs. Externalization. Externalization allows greater perception of both spatial depth and width and also perception of the position of the source outside of the head.

- **Front-back Confusion**

Cone of confusion restricted to only horizontal plane refers to widely-discussed phenomenon of *front-back confusion* [83], [84]. Identical ITD and ILD can correspond either to front or rear position of the sound source. The crucial parameter for ability to distinguish, whether the source is situated in front of the listener or behind the listener, is determined primarily by shadowing of high frequencies by pinna structure. This is important in static virtual positioning, when the subject can perceive the position according to *brightness* of the sound. However, this might become unreliable, when the listener can not compare the variants of front and back signal. In real listening conditions with real source, the position can be determined by tiny head movements [83], which are helpful for perception of the spatial arrangement of the acoustic scene[18]. This situation is illustrated in Fig. 2.29. In the initial position, two sources are placed in fixed location φ_{i11} for the first source and φ_{i21} for the second source. Since they are placed axially symmetric $\varphi_{i21} = 180° - \varphi_{i11}$. After a small head turn by φ_t both sources are perceived in a new positions; φ_{i12} for the first source and φ_{i22} for the second one. The final position is denoted by dark gray, the initial position is marked by light gray. For the case of front position φ_{i11}, after the head turn the source position shifts to the side (edge) position, while for back position φ_{i11} the source shifts to front-back axis. These shifts of the source position allows to determine the front/back half plane. This effect can be emulated by implementing head-tracking sensor to the positioning system. The prevalence of front-back confusion can help to rate particular positioning methods or their upgrades [85].

[18] Note that for people with no specific sight impairment, the majority of spatial information is delivered mostly by visual perception.

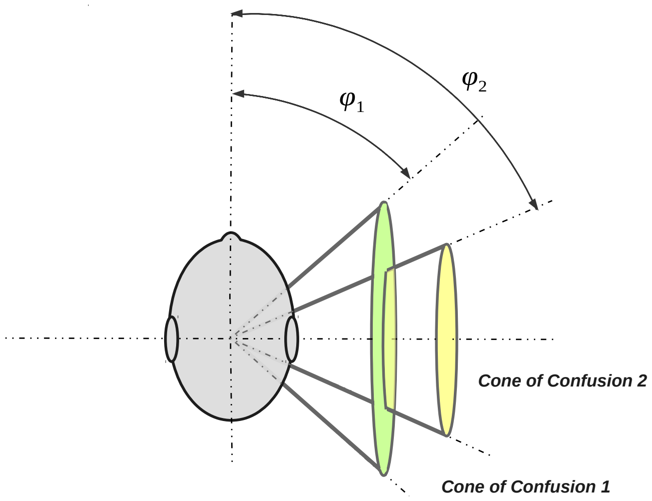

Figure 2.28: Cone of confusion. Cone of confusion determines positions with identical ITD and ILD. It is usually regarding a head approximated by a sphere.

Acquisition the rating scores requires a reliable method. Since the majority of spatial perception issues are subjective-dependent with particular variance among the listeners, a design of specific listening tests is needed. The following sub-section introduces several methods of listening tests utilized within this thesis.

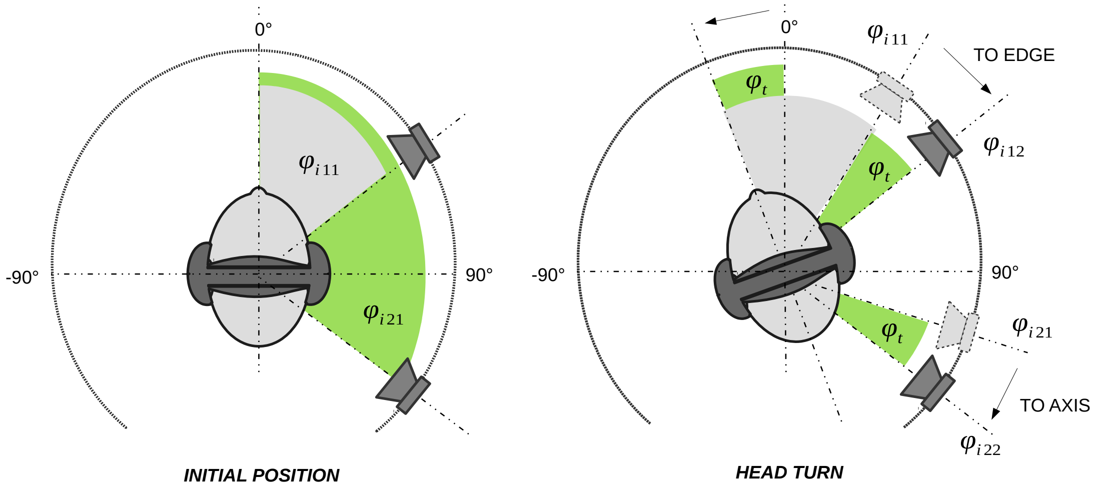

Figure 2.29: Front-back confusion. The source position changes with a head turn. Depending on the front or back position, the virtual source moves towards the *edge* or *axis* position.

2.6.1 Measuring Methods

In order to obtain valuable and correct results, it is necessary to focus as on the design of listening test, as on its performance, and also its evaluation. Composition of the test requires choosing an appropriate amount of stimuli with its variation of the explored parameter adequate to appropriate range; to preserve reliability and validity. When not designed well, the subject might be confused, misrepresent answers, not pay attention, or simply not understand the question. A neat summary of performance of listening tests can be found in [72]. This section provides a brief survey of methods for listening tests that are suitable to be applied to the field of virtual positioning. Most of them are employed within this thesis.

- **Paired comparison**

  In paired comparison method, the subject is presented by two stimuli and then asked to compare both in terms of specific attribute. There is no objective reference, since each of the two stimuli in the pair is referenced to the other. The paired comparison method is able to align the stimuli ascending according to perceived measure of a specific parameter. However, there is no objective quantification of *how much* the stimuli differ from each other. The advantage of this method lies in ability to check the consistency of the answers. For instance, when stimulus A is rated better than B and stimulus B better than C, then it is expected the subject will rate A better than C as well:

  $$A > B, B > C \quad \Rightarrow \quad A > C. \tag{2.24}$$

  Number of stimuli pairs within a group of size s is

  $$C_2(s) = \binom{s}{2} = \frac{1}{2}\left(s^2 - s\right). \tag{2.25}$$

  However, disadvantage is that the number of pairs rapidly increases with the number of stimuli to be compared.

- **Rank order**

  Ranking is one of the most intuitive methods for the subjects to understand the task. The subject is given a certain amount of stimuli and is asked to put them in order according to particular attribute. As in *paired comparison*, the results finally deliver only the order of stimuli. There is no information about level of quantity of the stimuli variance.

- **A / B / X**

  When the difference of both stimuli is only very slight, A/B/X test method shall be applied. This method discovers whether the difference of two stimuli is perceptible or not. This method is based on presenting three stimuli to the subject. Stimulus A is the original, stimulus B is the modified one, and stimulus X is always either

A or B. The subject is asked to find whether X is identical to A or B. Result with 100% score of this test refers the difference between A and B is well perceptible and 50% result refers the subject was guessing and there is no noticeable difference. The subject should be trained for the range of differences before the experiment starts.

- **Three-alternative forced choice (3AFC)**

  The 3AFC method is an alternative for mentioned A/B/X method. There are also 3 stimuli (A, B and C) presented to the subject . The subject is aware that two stimuli are identical and one differs. His task is to choose the different one. This method might give better results than A/B/X, when the difference between the stimuli is very slight, since there is only 33.3% of selecting the correct answer opposite of 50% in A/B/X method.

- **Two-alternative forced choice (TAFC or 2AFC)**

  In this method, the subject is asked to make a choice between two stimuli and decide which one prevails according to the desired parameter. The subject has to make the choice even when the difference is imperceptible. This method is similar to paired comparison, however, character of presented stimuli is adapting to answers of the subject. In the beginning of the test, the difference is at such level that the subject certainly detects the difference. Each time the subject gives a correct answer, the difference between both stimuli decreases. Each time the subject gives a wrong answer, the difference between both stimuli increases. The difference between examined parameter of the stimuli can be reduced to the limit of the JND. In such case there is a possibility of looping of increasing/decreasing the difference, therefore it is necessary to stop asking the questions after determined number of reversals and find the mean value.

- **Double-Blind Triple-Stimulus with Hidden Reference (BS.1116)**

  The subject is given 3 stimuli as in A/B/X test. One is the reference, one is identical to the reference and the last on differs. The subject is asked co complete two tasks. In the first one, which of the stimuli A and B is identical to the reference. The second task is to rate the level of the difference of the remaining stimuli by a five-point scale (5 - Imperceptible, $\cdots$, 1 - Annoying). If the subject correctly understands the question, at least one of the stimuli will have rating of 5. BS.1116 is more complex as to interpret as to evaluate than the A/B/X.

To gather the responses of subjects, a graphical user interface can be used, where the subjects adjust particular sliders or clicks at specific location. In terms of examining localization precision or offset, the subjects can simply choose the correct answer from a given set of positions in a questionnaire (paper or GUI) or use a laser pointer to determine the position while being recorded by a camera. The choice of a method for testing particular attributes of a virtually positioned sound and choice of way, how the subject responds, depends on the parameters to be explored and on the stimuli to be selected.

Chapter 3

Proposed Method: Positioning by Differential HRTF

The HRTF contains localization cues for a human listener, i.e. Interaural Time Difference (ITD), Interaural Level Differences (ILD) and spectral cues. To obtain virtually positioned sound by the HRTF method, a convolution of the original signal and appropriate HRIR pair is required. This two-channel processing may increase requirements for computational resources, when multiple sources are rendered simultaneously (e.g. in computer games, training assistive programs for visually impaired or for low-cost solutions). Many articles have already dealt with more effective measuring or rendering of the HRTFs. However, simplifying the positioning process focused on reduction of the computational resources is not well-explored issue yet. Therefore, there raise a need of algorithm with better spatial performance than the simple amplitude panning with less processing requirements than the HRTF. This chapter introduces a new approach to virtual sound source positioning primarily focused on reducing the computational costs, an experimental method called *Differential Head-Related Transfer Function* (DHRTF).

3.1 Method Background

The first pilot study by the authors has been presented in [A.6, A.5]. Assume that the common AP processing changes the amplitude ratio in both channels. The final perceived in-head position does not depend on the absolute amplitude of both signals, but on their difference expressed by the ILD. It can be also assumed that both signals are not approaching extreme high or low levels within the hearing dynamic range. When the HRTF positioning method is applied, separate HRTF filtering results in mutual differences in both channels and frequency-dependent ILD and ITD emerge. The principle of the Differential HRTF lies in introduction of the frequency dependent ILD and ITD to the stereo signal. Therefore, filtering by a pair of the HRTFs is reduced to a one-channel filtering, where the same inter-channel differences occurs in the positioned sound as when filtered by the HRTF. Only one channel is processed while the other one remains completely untouched.

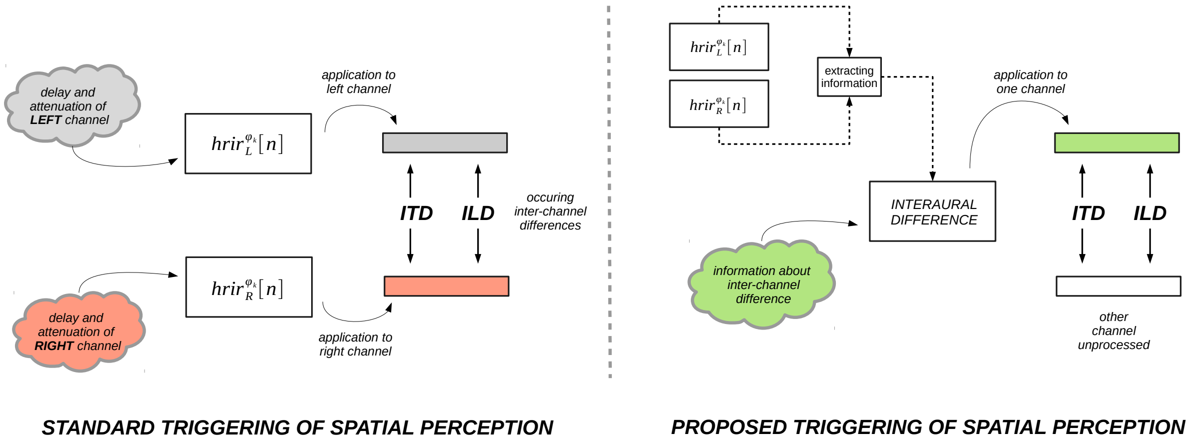
delay and attenuation of LEFT channel
$hrir_L^{\varphi_k}[n]$
application to left channel
ITD
ILD
occuring inter-channel differences
delay and attenuation of RIGHT channel
$hrir_R^{\varphi_k}[n]$
application to right channel
STANDARD TRIGGERING OF SPATIAL PERCEPTION
$hrir_L^{\varphi_k}[n]$
$hrir_R^{\varphi_k}[n]$
extracting information
application to one channel
INTERAURAL DIFFERENCE
information about inter-channel difference
ITD
ILD
other channel unprocessed
PROPOSED TRIGGERING OF SPATIAL PERCEPTION

Figure 3.1: HRTF versus DHRTF principle. HRTF (top) and DHRTF (bottom) positioning methods are compared here. In the HRTF method, both L and R channels are processed in parallel. In the DHRTF method, information about time and level differences is extracted from the HRTF pair and applied to only one channel of the stereo signal.

Taking into account that source localization in the horizontal plane is more important than in the median plane [9, 74, 50], the most significant parameters regard the ILD and the ITD. In general, the final spatial perception of the real (i.e. wide-band mostly) sound sources does not depend absolutely on the particular spectral content [9]. Therefore, finding appropriate ILD and ITD is crucial for success of the spatial illusion within the horizontal plane.

As demonstrated in Fig. 3.1 (a), the common HRTF positioning is based on affecting each of the stereo channels by appropriate HRIR, which stores the information of sound attenuation and time delay. When the HRTF positioning method is applied, separate HRTF filter application results in mutual differences in both channels and frequency-dependent ILD and ITD. The new proposed approach of virtual positioning is based on extraction of the proportions of the ILD and the ITD for a given pair of HRIRs and its application to the stereo signal. Generally, within the horizontal plane, there is always one channel delayed and attenuated in respect to the other. Therefore, ITD and ILD can be extracted from an existing HRTF pair and these differences can be then applied to one of the channel, as demonstrated in Fig. 3.1 (b). The resulting signal will dispose by the same localization cues, however, achieved by processing only one channel. Therefore, only one channel is processed while the other remains completely untouched.

The above-mentioned facts initialized an idea of utilizing the information from the HRTF and simplifying the positioning process. The primary objective is to preserve the fidelity of the virtual space related to the frequency dependence of the ILD and ITD. This resulted in the concept of *Differential Head-Related Transfer Function* (DHRTF).

3.2 Differential Head-Related Transfer Function

Definition of the Differential HRTF requires to introduce several more terms. As equivalent of ITD, also *Interaural Phase Difference* (IPD) can be defined. The phase information of particular frequency components is contained in the phase characteristics of the complex HRTF. Therefore, IPD can be obtained as difference of both phase characteristics of the HRTF pair, which can be expressed as

$$\mathrm{IPD}^{\varphi}(\omega) = \psi_c(\omega) - \psi_i(\omega) = \arg\Big(\mathrm{FT}\{hrir_c^{\varphi}(t)\}\Big) - \arg\Big(\mathrm{FT}\{hrir_i^{\varphi}(t)\}\Big), \tag{3.1}$$

where $\psi_c(\omega)$ and $\psi_i(\omega)$ denotes the phase of the contra-lateral and the ipsi-lateral channels, respectively, and $\arg(\cdot)$ extracts the phase from a complex number. Notice also content in Eq. (2.1). The Differential HRTF is defined as the ratio of the contra-lateral and ipsi-lateral HRTFs. The DHRTF can be considered as a transfer function, which magnitude is actually frequency-dependent ILD and the phase corresponds to frequency-dependent IPD, as stated below

$$\mathrm{DHRTF}^{\varphi}(\omega) = \frac{|\mathrm{HRTF}_c^{\varphi}(\omega)|}{|\mathrm{HRTF}_i^{\varphi}(\omega)|} \cdot \mathrm{e}^{j(\psi_c(\omega)-\psi_i(\omega))} = \mathrm{ILD}^{\varphi}(\omega) \cdot \mathrm{e}^{j\cdot \mathrm{IPD}(\omega)}, \tag{3.2}$$

where indices c and i stand for the contra-lateral and ipsi-lateral ear, respectively, ψ is the phase spectrum of a standard HRTF corresponding to particular ear. Application of the inverse Fourier transform on the DHRTF produces its equivalent in time domain *Differential HRIR* (dHRIR)[1], which is used in the signal processing (specified bellow). Correspondence to the ipsi- or contra-lateral channel is defined as follows

$$\mathrm{HRTF}_c^{\varphi}(\omega) = \begin{cases} \mathrm{HRTF}_L^{\varphi}(\omega) & \text{for} \quad \varphi \in (0, 180) \\ \mathrm{HRTF}_R^{\varphi}(\omega) & \text{for} \quad \varphi \in (-180, 0) \end{cases} \tag{3.3}$$

$$\mathrm{HRTF}_i^{\varphi}(\omega) = \begin{cases} \mathrm{HRTF}_L^{\varphi}(\omega) & \text{for} \quad \varphi \in (-180, 0) \\ \mathrm{HRTF}_R^{\varphi}(\omega) & \text{for} \quad \varphi \in (0, 180). \end{cases} \tag{3.4}$$

The reason for this strict division is explained in subsection 3.2.1 and ensures proper function of the positioning algorithm. The procedure described above describes the extraction of information about spectral and time differences between both stereo channels in the real listening conditions. For demonstration of the process, see Fig. 3.2. The HRTF pair in (a) shows a unique gain of both transfer functions. Typical weak-dependent lower frequencies (under 1 kHz) and several specific peaks and notches at higher frequencies (around 10 kHz) are visible. The Differential HRTF derived from this pair is shown in (b) of the same figure. This DHRTF may be also understood as the frequency dependent ILD for particular position ($\vartheta = 70^\circ$ and $\vartheta = 0^\circ$). Notice that the surface beneath the curve in (b) equals the surface circumscribed by the two HRTF curves in (a).

[1] For better orientation within the terms, time domain is described with small letter d (dHRIR), while the frequency domain is written with capital D.

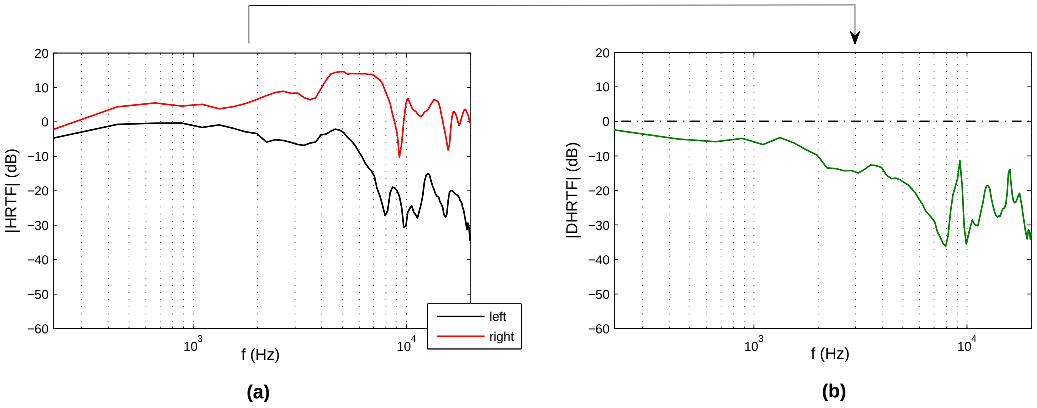

Figure 3.2: DHRTF derived from the HRTF pair. An example of the HRTF pair (a) and corresponding Differential HRTF (b) for $\varphi = 70°$. The magnitude of the DHRTF is actually a frequency-dependent ILD with typical variance at high frequencies.

For the purposes of virtual positioning, the DHRTF needs to be transformed back to the time domain in order to perform linear convolution effectively (see Section 4.2). Application of the inverse Fourier transform on the DHRTF results in the *Differential Head Related Impulse Response* (dHRIR) that is used for the implementation

$$dhrir^{\varphi}(t) = \mathrm{FT}^{-1}\Big\{\mathrm{DHRTF}^{\varphi}(\omega)\Big\}. \tag{3.5}$$

The process of obtaining a set of dHRIRs from a set of HRIRs is graphically summarized by block diagram in Fig. 3.3.

The theoretical concept of virtual positioning by the *Differential Head-Related Transfer Function* (DHRTF) has been already introduced in [A.13], [A.6], and [A.5] by the author. The concept of employing the ratio of the contra- and ipsi-lateral HRTFs being referred to as *Interaural Transfer Function* (ITF, IATF) has been previously used in several applications. The ITF was employed for cross-talk cancellation in [40], for modeling of the contra-lateral HRTF from a measured ipsi-lateral [79], or for low-order approximation of the contra-lateral HRTF [86]. However, it has never been used in a concept of direct headphone-based virtual positioning. The author use designation *Differential* HRTF to underline employment of the ITF as a one-channel positioning method (*differential* refers to difference of the two HRTFs in the logarithmic scale).

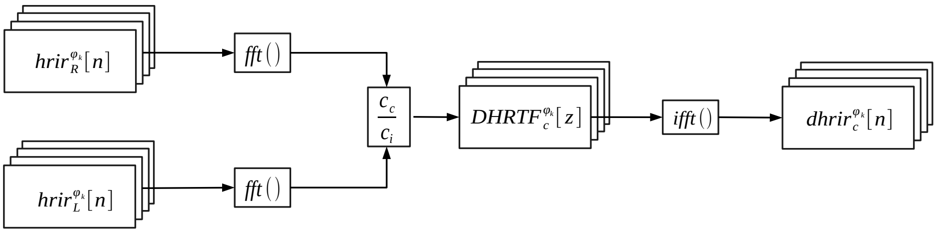

Figure 3.3: Obtaining DHRTF. A block diagram of obtaining the Differential HRIR set from an already existing set of HRIRs.

3.2.1 One-dimensional Demonstration

The equations above define the general concept of the DHRTF. However, in order to comprehend the extraction of the ITD and the ILD from HRIR pair, a brief demonstration is advisable. This example introduces behavior of dHRIR for one fixed position in discrete time. For simplicity, assume having HRIR for each ear sampled in discrete time with length of 100 samples. The impulse response itself is represented by only one non-zero sample (i.e. the frequency response is frequency-independent) of specific amplitude. In this case of demonstration the responses are described as follows

$$hrir_L^{\varphi}[n] = \begin{cases} 0.5 & \text{for } n = 70 \\ 0 & \text{elsewhere} \end{cases} \tag{3.6}$$

$$hrir_R^{\varphi}[n] = \begin{cases} 2 & \text{for } n = 50 \\ 0 & \text{elsewhere .} \end{cases} \tag{3.7}$$

The standalone sample represents the gain of the HRTF by its value and the time delay of the channel by its position in the impulse response. For our purposes, the right channel is denoted as ipsi-lateral, therefore its impulse response starts earlier and with higher energy. This case is demonstrated in Fig. 3.4, where h_R and h_L represent the standalone samples of the impulse response with different amplitudes (2 and 0.5 respectively) and with ITD length of 20 samples. By performing Eq. (3.2) and Eq. (3.5) a correct differential impulse response h_{D1} is obtained. Its position from the beginning of the response corresponds to the ITD and its amplitude of 0.25 corresponds to the amplitude ratio of both HRIRs. The definition of the DHRTF strictly requires ratio of the contra-lateral channel to the ipsi-lateral. Otherwise, it would result in h_{D2} which is not applicable in the virtual positioning, since the ITD does not correspond to the actual value.

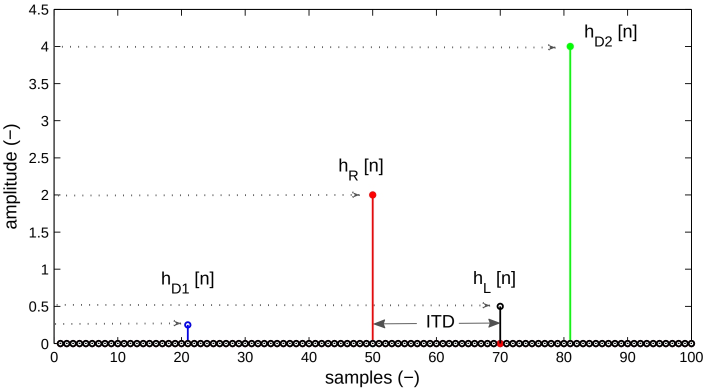

Figure 3.4: 1-D DHRTF demonstration. Demonstration of $dhrir[n]$ obtained from simplified responses represented by $hrir_L[n]$ and $hrir_R[n]$. The correct difference impulse response is h_{D1}. The latter h_{D2} is not applicable.

3.2.2 Two-dimensional Demonstration

The single case demonstration introduces what happens with the samples of the differential impulse response. In order to demonstrate behavior of dHRIR in the whole set across the positions (i.e. for $\varphi \in (-180, +180)$) a trivial model describing general HRIR behavior was implemented. This model also takes into account one-sample HRIR, moreover it is extended that it covers the dependence of the amplitude and the time shift on position. The highest gain and the earliest onset is observable for ±90° (corresponds to +90° and +270° for the positive interval) positions in dependence on particular ear and vice versa the lowest gain and the latest onset is for the same positions for the other ear. The time shift (ITD) here was modeled by harmonic function with linear decrease and the onset starts between sample 10 and sample 50. The simple set of the one-sample HRIRs is demonstrated in Fig. 3.5; notice the same trend of amplitude of the first peak in Fig. 2.21. The samples can be calculated from equation

$$hrir^{\varphi}_{Lm}[n] = \begin{cases} 0.55 + 0.45 \cdot \sin(\frac{\varphi\pi}{180}) & \text{for } 30 - \lfloor 20 \cdot \sin(\frac{\varphi\pi}{180}) \rfloor \\ 0 & \text{elsewhere} \end{cases} \tag{3.8}$$

for the left ear and proportionally as

$$hrir^{\varphi}_{Rm}[n] = \begin{cases} 0.55 - 0.45 \cdot \sin(\frac{\varphi\pi}{180}) & \text{for } 30 - \lfloor 20 \cdot \sin(\frac{\varphi\pi}{180}) \rfloor \\ 0 & \text{elsewhere} \end{cases} \tag{3.9}$$

for the right ear. Angle φ denotes the position in the horizontal plane in degrees. After application of the algorithm defined by Eq. (3.2) and Eq. (3.5) a set of dHRIRs is created. The set resulting from the simple pair of HRIRs introduced above is shown in Fig. 3.6.

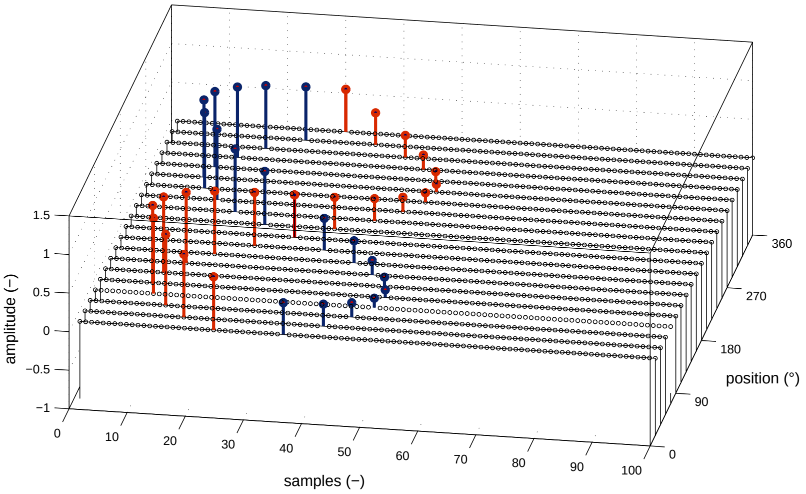

Figure 3.5: Simple HRIR set demonstration. Visualization of the modeled simple one-sample set of $hrir_L[n]$ (blue) and $hrir_R[n]$ (red). Typical shape of particular time shifts creates a *wave* in the visualization.

The time shift corresponds to the time shift between the left and the right HRIRs with its maximum of 40 samples. Note that the inter-channel delay is always in positive values since the ear channels switch their contra- and ipsi-lateral role for the particular positions. No time difference is observable for the positions on the front-back axis ($\varphi = 0°$ and $\varphi = 180°$). The highest amplitude difference occurs (as expected) for positions $\varphi = 90°$ and $\varphi = 270°$. As defined before, the dHRIR (DHRTF) contains the information about the time shift and the gain attenuation of the contra-lateral channel. For clarity and comparison, see actual measured values of both HRIR onsets and resulting ITD obtained from a set of HRIRs measured on acoustic manikin available in [17] (red and black, dotted), which are shown in Fig. 3.7. The time information was obtained by thresholding of the impulse response energy. The ITD (i.e. the delay of the dHRIR) is depicted by the dashed line, showing orderly structure of two triangles. In the standard sampling rate of 44.11 kHz, the maximum ITD corresponds to approximately 30-34 samples in dependence on the head proportions. Compare the response time delay and resulting actual ITD values in the Fig. 3.7 with the Fig. 3.6.

Fourier transform of the simple dHRIR set from Fig. 3.6 results in series of flat (frequency independent) transfer functions of the DHRTF with position-dependent magnitude, as demonstrated in Fig. 3.8. These transfer functions denotes attenuation of the contra-lateral ear. Maximum of the magnitude is 0 dB, which corresponds to unity delta function response for $\varphi = 0°$ and $\varphi = 180°$. In real situations the gain of the contra-lateral ear should not exceed this level. However, this assumption is not always fulfilled in real cases,

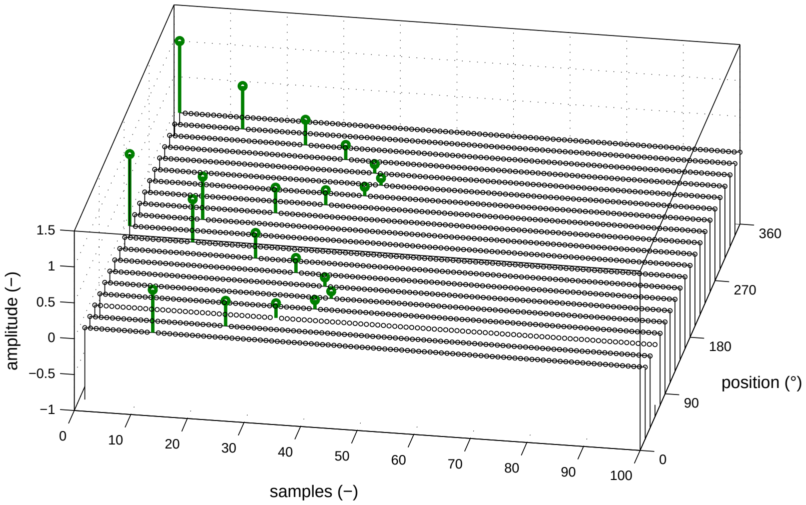

Figure 3.6: Simple DHRIR set demonstration. Visualization of a set of $dhrir[n]$ resulting from a set of $hrir_L[n]$ and $hrir_R[n]$ in the previous Fig. 3.5. The ITD varies between 0 and 40 samples.

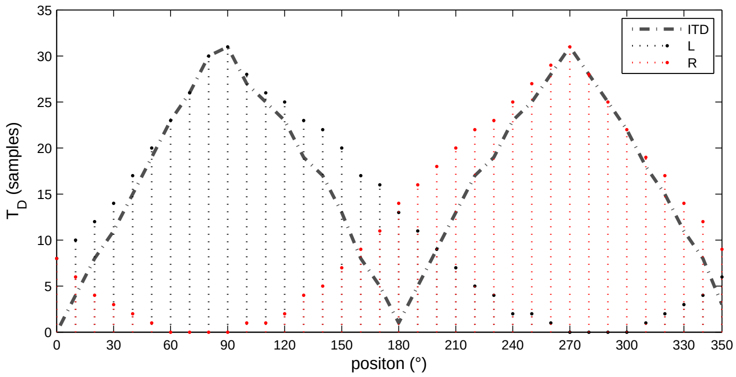

Figure 3.7: ITD from HRTF set. The Interaural Time Delay extracted form a real HRTF set by detecting the signal energy onset of each impulse response. The dotted black and red lines correspond to the time shift of left and right HRIRs, respectively. Dashed triangles represent the resulting ITD.

as shown in the further text. Note the transfer functions are flat due to only one sample present in respective impulse responses. This representation of HRIRs would corresponds to the amplitude panning method extended by ITD implementation (the ITD is hidden in phase characteristics of the transfer function). Frequency dependence would occur when more samples were added to the corresponding impulse responses.

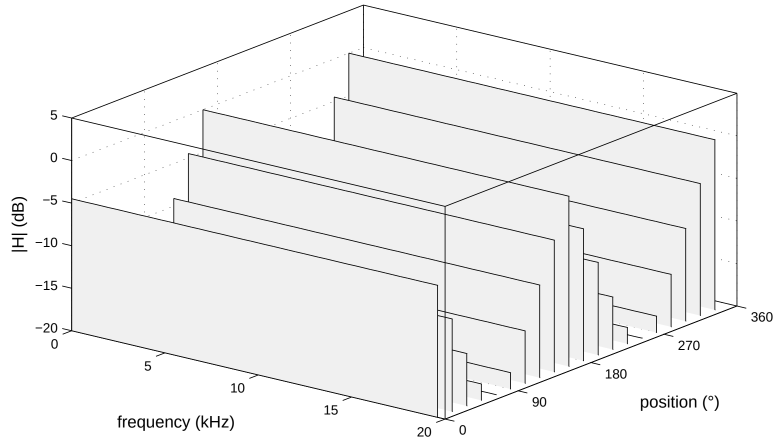

Figure 3.8: Simple DHRTF set demonstration. Differential HRTF resulting from a set of $hrir_L[n]$ and $hrir_R[n]$ in Fig. 3.5. Various amplitudes of one-sample $dhrir[n]$ in Fig. 3.6 results in a flat transfer function with various gain.

3.3 Virtual Positioning by the DHRTF

The essence of the DHRTF algorithm lies in introducing the frequency dependent ILD and ITD to the stereo signal not by separate filtering of both channels of the binaural signal (as is in the HRTF method), but by filtering only one channel in a way that the same inter-channel differences will occur in the stereo sound, as when the HRTF method is applied. Therefore, only one channel is processed, while the other one remains completely untouched. The difference between both algorithms is demonstrated in Fig. 3.9. The algorithm based on the DHRTF processing is similar to the standard HRTF positioning as it employs convolution of the signal and the filter response. The channels are now not separated to the *left* and *right*, but to the *contra-lateral* and *ipsi-lateral* instead. Obtaining of dHRIR is described by Eq. (3.5). In terms of signal processing, the response of the ipsi-lateral channel can be defined as standalone discrete delta function on the first index of the signal vector [2]

$$dhrir_i^{\varphi}[n] = \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.10}$$

[2] There exists various indexing among programming languages. For purposes of this thesis, the standard C/C++ concept was chosen for description. Therefore, the first sample of the signal vector has an index of 0. Notice that some examples further in the text introduce segments of Matlab code, where the first element of a vector is indexed by 1.

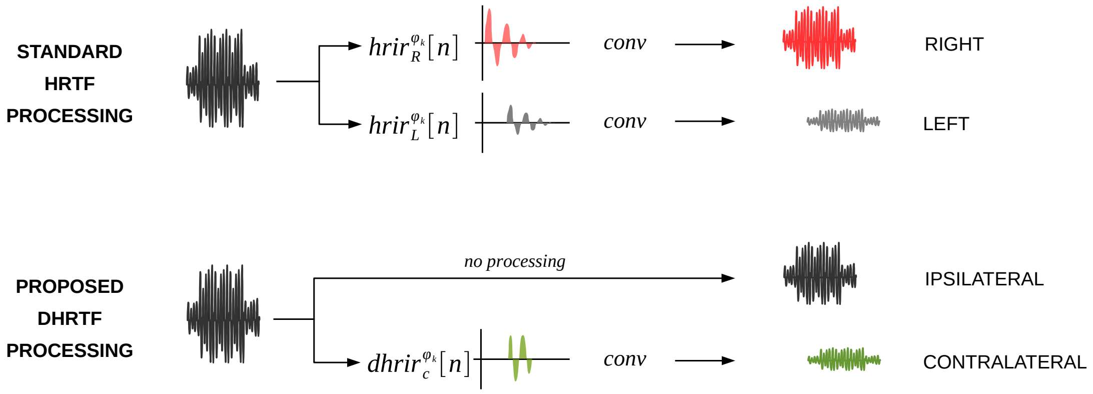

Figure 3.9: HRTF vs DHRTF processing. The concept of standard virtual source positioning by the HRTF and positioning by the Differential HRTF is compared. Both methods result in the same amplitude ratio and the same inter-channel time shift.

Since convolution of the delta function $\delta[0]$ and the signal $x[n]$ returns the original signal $x[n]$, static positioning by DHRTF can be summarized as

$$y_c^{\varphi}[n] = dhrir_c^{\varphi}[n]{*}x_m[n] \tag{3.11}$$

$$y_i^{\varphi}[n] = x_m[n], \tag{3.12}$$

where $y_c^{\varphi}[n]$ and $y_i^{\varphi}[n]$ denotes the contra- and the ipsi-lateral channels, respectively, and $x_m[n]$ represents the original monaural sound. Compare these equations to the corresponding equation of the HRTF method, Eq. (2.10). Therefore, whenever dHRIR processing is applied, always only the contra-lateral channel is attenuated and delayed with respect to the original $x[k]$, which also takes place in the ipsi-lateral channel of the signal. Block diagram representing the signal flow in the positioning process is shown in Fig. 3.10.

Since the information about monaural spectral cues that is essential for sound localization in median plane [74, 25] is lost or heavily distorted during the processing (only ITD and ILD remain), the DHRTF is limited only to horizontal positioning, as the AP is. This is the weakest attribute of the DHRTF positioning method. However, in terms of practical application, horizontal motion (consider virtual scene of walking person) usually prevails over a vertical motion.

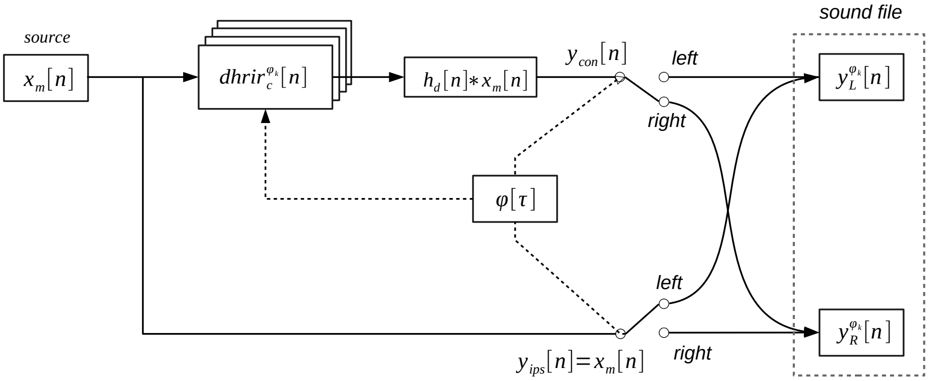

Figure 3.10: DHRTF processing scheme. Block diagram of virtual sound source positioning by the DHRTF. Compare with standard HRTF positioning in Fig. 2.15 is depicted. The processed signal designated for the conta-lateral channel is routed to the left or right channel of the stereo file, according to the source position.

3.4 DHRTF set from a real measured HRTF

Resulting from the previous statements, DHRTF is primarily dependent on the original set of the HRTFs, from which it is derived. As already introduced in Section 2.5, several approaches are possible to get an HRTF set. In case of measured HRTF, equalization of the measuring chain is required for better performance. The principle of the DHRTF allows to effectively eliminate these attributes. As states in Eq. (2.16), the measured HRTF is subject to influence of particular transfer functions involved in the measuring chain, i.e. the microphone and the loudspeaker characteristics, and the room response. By combination of Eq. (2.16) and Eq. (3.2) and by modification of the resulting equation, we can obtain

$$\begin{aligned} dhrir_c^{\varphi}[n] &= \mathrm{FT}^{-1}\left\{\frac{\mathrm{HRTF}_{\xi_1}^{\varphi}[\Omega]}{\mathrm{HRTF}_{\xi_2}^{\varphi}[\Omega]}\right\} \\ &= \mathrm{FT}^{-1}\left\{\frac{H_{me,\xi_1}^{\varphi}[\Omega]}{H_{r,\xi_1}^{\varphi}[\Omega]\cdot H_l[\Omega]\cdot H_m[\Omega]}\cdot\frac{H_{r,\xi_2}^{\varphi}[\Omega]\cdot H_l[\Omega]\cdot H_m[\Omega]}{H_{me,\xi_2}^{\varphi}[\Omega]}\right\} \\ &= \mathrm{FT}^{-1}\left\{\frac{H_{me,\xi_1}^{\varphi}[\Omega]\cdot H_{r,\xi_2}^{\varphi}[\Omega]}{H_{me,\xi_2}^{\varphi}[\Omega]\cdot H_{r,\xi_1}^{\varphi}[\Omega]}\right\} \\ &\cong \mathrm{FT}^{-1}\left\{\frac{H_{me,\xi_1}^{\varphi}[\Omega]}{H_{me,\xi_2}^{\varphi}[\Omega]}\right\}, \end{aligned} \tag{3.13}$$

where $H_{me,\xi_1}^{\varphi}[\Omega]$ represents the measured HRTF, $H_m[\Omega]$ stands for the transfer function of the measuring microphone, $H_l[\Omega]$ is the transfer function of the loudspeaker and $H_{r,\xi_1}^{\varphi}[\Omega]$

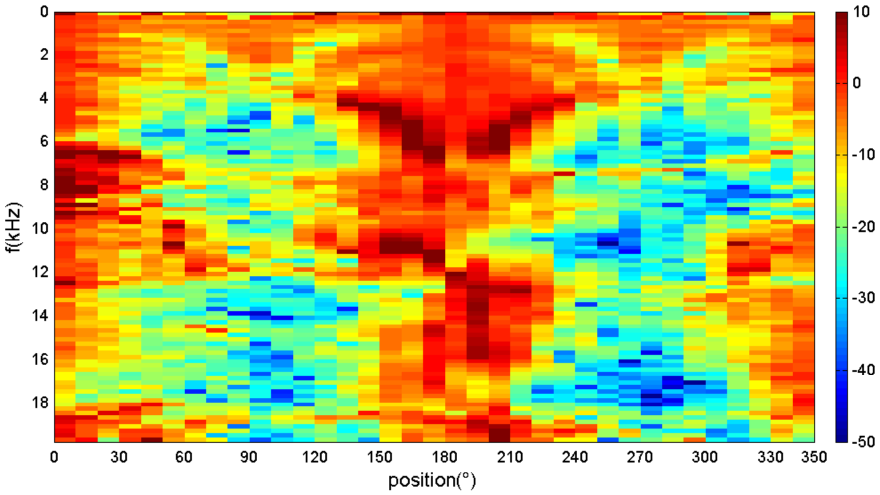

Figure 3.11: Heat map of DHRTF magnitude. The DHRTF set corresponding to a real measured HRTF set previously introduced in Fig. 2.23 is demonstrated.

represents the room transfer function. The approach assumes all the transfer functions to be identical for both ears. This may not be fulfilled for the room impulse response perfectly, since both ear positions may receive different sum of reflection. However, the HRTF is mostly measured in an anechoic room, thus no reflection occurs. Other way to remove the room impulse response it to simply limit the length of the RIR, as states in Eq. (2.17). The DHRTF derived form a real measured set introduced in Fig. 2.23 is presented in Fig. 3.11. The frequency axis (y) in the figure is kept in linear scale for better observation of its behavior. The frequency-dependent features are practically symmetrical with their highest attenuation for side positions (90° and 270°) The ILD occurs up to 50 dB. Note that the DHRTF contains switching of the contra- and the ipsa- positions of the ears, thus the difference to the other ear is negative.

3.5 Issues in positioning by Differential HRTF

In specific HRTF pairs, an unexpected phenomenon occurs. In unfavorable constellation the attenuation of the ipsi-lateral channel may be greater than in the contra-lateral (against expectation) for particular frequencies, having a character of a narrow-band notch. From the definition of the DHRTF, the same ILD is present here. Therefore, the increased gain is caused by presence of sharp peak in the DHRTF (spike) exceeding the level of 0 dB. Since the property of the DHRTF lies in frequency-dependent attenuation and time delay of the contra-lateral channel, positive value of the DHRTF results in boosting the specific band in this channel channel.

Well observable, Fig. 3.11 also shows dark red areas, which indicates that level of the magnitude exceeds the edge of 0 dB at particular frequencies. This phenomenon does not

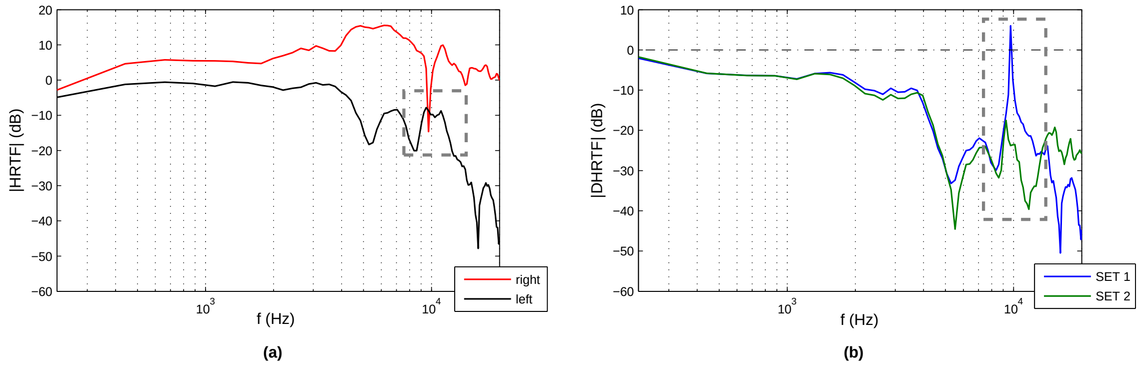

Figure 3.12: Origin of the DHRTF artifacts. Magnitude of the left and right HRTF for $\varphi = 85°$ and $\vartheta = 0°$ (a) is shown. When a notch attenuation of the HRTF for the ipsi-lateral ear crosses the HRTF of the contra-lateral, a spike-like peak in the DHRTF occurs (b). Spectral peaks and notches vary uniquely in accordance with the source position and selected HRTF set. The green line in (b) represents a DHRTF without a spike for the same location for another subject.

correspond to the previously stated assumption that the gain of the contra-lateral channel is always lower of the ipsi-lateral. The phenomenon was named Negative Inter-aural Level Difference (NILD) [A.2], [A.1]. The NILD occurs especially within the DHRTF derived from a real measured set. The origin of its occurrence results from a unique constellation of the HRTFs in a pair. As shown in Fig. 3.12 (a), the transfer function of the ipsi-lateral channel can have a spectral notch, which crosses the transfer function of the contra-lateral channel, i.e. the contra-lateral gain is lower than the ipsi-lateral for specific frequency band. This results in spectral spike presence in the DHRTF magnitude, as shown in Fig. 3.12 (b), blue line. Panel (b) contains DHRTF derived from two different sets of the HRTF corresponding to the same position. The blue line refers to the inappropriate one. However, the presence of the spike in the DHRTF is neither determined for specific spectral bands, nor for specific positions and appears to occur chaotically. For instance, green line in the figure represents the DHRTF for the same position obtained from another subject. No spectral spike is observable here. Its occurrence is different among various DHRTF sets.

Figure 3.13 shows the extracted NILD spikes in one DHRTF set already shown in Fig. 3.11. This set of the DHRTF is heavily distorted by the unwanted spectral features. As can be observed, the NILD occurs mostly in form of narrow spikes. However even wider frequency bands can appear in a particular DHRTFs, especially around frontal axis, i.e. positions $\varphi = 0°$ and $\varphi = 180°$. Due to the principle of the method, the artifacts are generally likely to occur around these positions. The Negative ILD may cause noticeable disturbing artifacts. Perception effect of the mentioned spectral spike is an unwanted pure tone character disturbance in the contra-lateral channel. This undesirable phenomenon

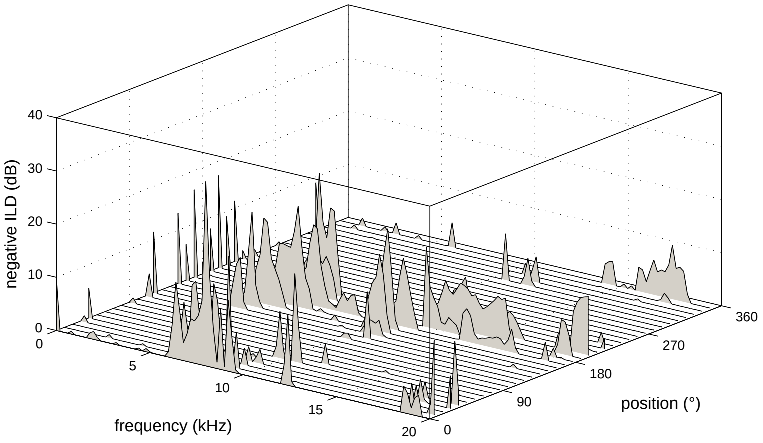

Figure 3.13: Negative ILD in the DHRTF. Spectral spikes of the Negative ILD ($|H_d[\Omega]| > 0$) extracted from the DHRTF introduced in Fig. 3.11 are depicted here.

can be avoided either by selection of appropriate DHRTF set (which actually limits personalization) or by adjusting of already selected individual set of the DHRTF (by filtering of spectral features or by spectral limiting/compression). This adjustment is considered to be performed prior to practical use.

More comprehensive analysis of the artifacts and their elimination (by employing spectral limitation and low-pass filtering for the DHRTF spectrum) is proposed and discussed as a part of the main results in the next chapter.

Chapter 4

Specific Aspects of DHRTF Method

This chapter summarizes the most important results obtained during the author's research. The text is divided into three sections, where each was used as a basis for particular publication.

Section 4.1 - *Comparison of Positioning Methods* analyzes the performance of the DHRTF within the horizontal plane. The DHRTF positioning method is compared to two other common methods: amplitude panning and HRTF processing. Results of theoretical comparison and quality assessment of the methods by subjective listening tests are presented here. The tests focus on distinctive aspects of the positioning methods: spatial impression, timbre affection and loudness fluctuations. The results show that the DHRTF positioning method is applicable with very promising performance; it avoids perceptible channel coloration that occurs within the HRTF method, and it delivers spatial impression more successfully than the simple amplitude panning method. The summary of this research was published in [A.2]

Section 4.2 - *Implementation of Moving Source* focuses on signal processing aspects of dynamic virtual sound source positioning. Rendering of the signal is based on application of filtering with varying impulse response of the filter (i.e. switching the HRTFs). This algorithm may be in its essence implemented by several methods, where each approach provides slightly different output signal resulting in variance of perception of spatial fidelity. The main difference results from a distinct output of linear and circular convolution usually implemented in such processing. Particular variants are described, discussed and compared by subjective listening test with the positioned signals. This material was published as an article in [A.4].

Section 4.3 - *Perception Artifacts in the DHRTF Method* presents specific methods for reducing artifacts, which may occur within virtual sound source positioning method employing the DHRTF. Three selected methods based on limitation and/or smoothening of the DHRTF magnitude by a low-pass filter are introduced here and compared by paired comparison listening tests and by objective assessment method. The tests showed significant decrease of the artifact occurrence with the best results for a method based on limiting the DHRTF magnitude and its smoothening by a moving average convolution kernel. The results were published in [A.1].

4.1 Comparison of Positioning Methods

In order to verify efficiency and practical usability of the proposed algorithm, the method of DHRTF was compared to standard approaches - *amplitude panning* and *HRTF processing* introduced in Chapter 2. In this section, the *Differential* HRTF positioning method developed by the author is compared to the AP and the HRTF in terms of quality of the rendered auditory space. Although it is very common to investigate primarily precision of a positioning method [50, 64, 87], this section focuses on particular aspects of perception of the virtual auditory environment; depth of the presented space, changes in timbre and fluctuations in loudness.

The principles of the introduced positioning algorithms are available in Sec. 2.2.1 for the AP, Section 2.3 for the HRTF, and Section 3.2 for the DHRTF). The next Subsection *Objective Comparison* reveals the objective differences between the particular positioning methods presenting their channel transfer function and position-dependent channel gain. The design and organization of the listening tests for assessing the methods is introduced later in *Subjective Comparison.* The consequent results are analyzed at the end of this section.

4.1.1 Objective Comparison of Positioning Methods

The three positioning methods described were examined for specific features. Energy of the channel response in dependence on azimuth (*gain curves*) and analysis of features within the direction-dependent channel transfer function along 360-degree radius in horizontal plane were focused. An example of outputs of the particular positioning methods for several periods of sine wave (*sine burst*) input is shown for demonstration.

4.1.1.1 Position-dependent Gain

In order to examine the total gain of *amplitude panning*, combination of Eq. (2.3) and Eq. (2.4) is used for the model. The following Eq. (4.1) and Eq. (4.2) give the analytical solution for computing appropriate position-dependent gains for left and right channel while preserving constant total energy of both channels.

$$g_L(\varphi) = \frac{1 - \sin(\varphi)}{\sqrt{2 \cdot \left(1 + \sin^2(\varphi)\right)}}, \tag{4.1}$$

$$g_R(\varphi) = \frac{1 + \sin(\varphi)}{\sqrt{2 \cdot \left(1 + \sin^2(\varphi)\right)}}. \tag{4.2}$$

where $g_{\mathrm{L}}(\varphi)$ and $g_{\mathrm{R}}(\varphi)$ refer to the respective channel gains and φ corresponds to the source angle position in the horizontal plane. As follows from the equations, gains for positions $\varphi \pm 90^\circ$ result in value of 0 for the contra-lateral ear. In real conditions this gain is nonzero. Maximum gain is reached for $\pm 90^\circ$ for the ipsi-lateral ear. Gains for front and

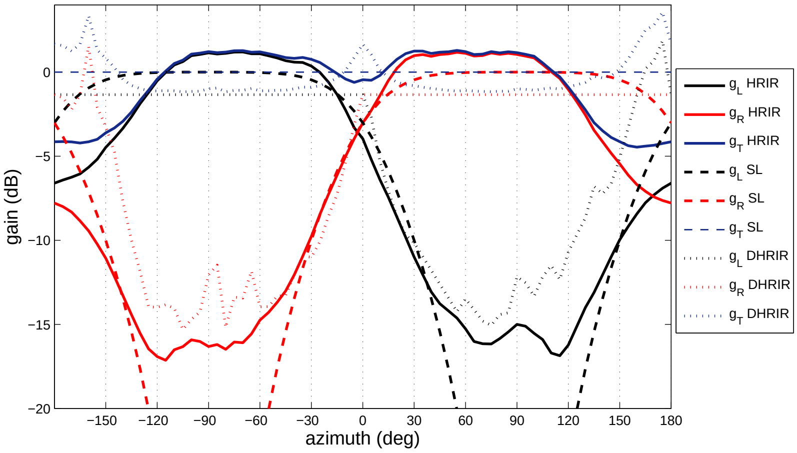

Figure 4.1: Gains of AP, HRTF and DHRTF. Logarithmic expressions of channel gains are shown here for three compared methods. Gains of the left, g_{L} and right, g_{R}, channels are shown together with the total gain g_{T}=g_{L}+g_{R}. Solid lines show HRIR gains of the HRTF, dashed lines show the sine law gains of the AP, and dotted lines show the DHRIR gain of the DHRTF method.

back half-plane are inherently symmetrical with no frequency dependence. Therefore, there is no option to distinguish between the two half-planes, front and back, from the signal. The varying channel gains for AP come out of analytical model solution. However, the HRTF positioning uses discrete-time impulse responses, thus the channel gain (i.e energy of the channel response) has to be obtained as follows:

$$g_{\xi}[\varphi] = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(hrir_{\xi}[\varphi, n]\right)^2}, \tag{4.3}$$

where HRIR represents impulse response of length N corresponding to particular azimuth φ. Gain for the ipsi-lateral channel of the DHRTF was computed according to Eq. (4.3). For this purpose, impulse response of the ipsi-lateral gain was considered as discrete Dirac pulse with unity gain. Resulting gain curves of the three methods are shown Fig. 4.1. In the figure, all the gain curves are shifted within the vertical axis in a way that the mean value of the total gain corresponding to the particular method (blue lines) is aligned to the value of 0 dB. An expectation of greater gain of the ipsi-lateral channel is reached in all the methods. However, there are several specific features in each method. As mentioned above, large attenuation of the contra-lateral channel of the AP around the side positions in the horizontal plane ($\varphi = \pm 90^{\circ}$) does not correspond to the actual gain derived from the HRTF. In order to preserve details, the vertical axis is limited to value of -20 dB; however attenuation for $\varphi = \pm 90^{\circ}$ reaches infinity for *panning* at these positions. The greater ILD of the AP might be useful for compensation of the missing ITD for creating better stereo impression. However, this effect causes void and unnatural character of the virtually positioned sound placed at these positions when listened by headphones.

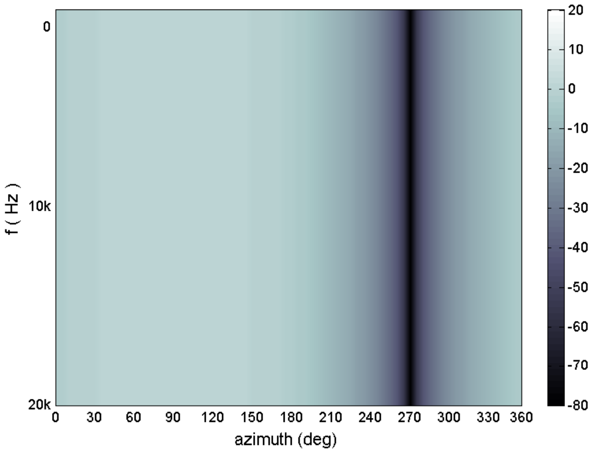

Figure 4.2: AP heat map. Behavior of transfer functions along the azimuth range is shown for the AP. Gain in dB corresponds to a particular position. The magnitude of the transfer function remains constant under the entire frequency range.

Unlike the *amplitude panning* method, the HRTF shows different course for some particular positions, even though the same trend of a rising gain for the ipsi-lateral channel of the gain curve is preserved. The most significant is the variation of the total gain. Notice also different total gain corresponding to front and back source positions. This phenomenon results from shadowing effects of pinna structure for back source positions at high frequencies. Another significant feature is a non-zero gain for side position of the contra-lateral ear. This feature has important role in natural sounding of the processed stimuli. In open space listening, the contra-lateral gain is reduced approximately by only 18 dB to the ipsi-lateral gain, as shows the HRTF method.

This behavior is quite well followed also by the DHRTF method. Another important feature is apparent for the central positions (around $\varphi = 0°$). While the total gain of the HRTF slightly decreases for about 2 dB compared to the maximal value at the edge position, the total gain of the DHRTF increases by similar amount at this position. Considerable increase of gain around the front and back positions ($\varphi \in \{0°, 180°\}$) results from occurrence of the negative ILD (see Sec. 3.5 and Sec. 4.3).

4.1.1.2 Position-Dependent Transfer Function

For the purposes of consistent comparison among the described methods, see heat map in Fig. 4.2 for demonstration of constant transfer function in y-axis. The heat map represents the magnitude of the right channel. The *amplitude panning* also lacks any time shift between the two channels, therefore the ITD is not present. This fact contributes to lack of *externalization* of the sound source (the sound source is perceived inside the head) and

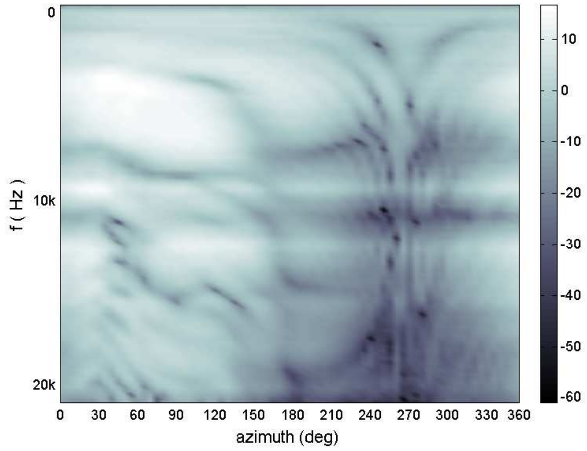

Figure 4.3: HRTF heat map. The HRTF of the right channel is shown here as a heat map for 360 degrees of the horizontal plane (φ is sampled by 5° step, y-axis shows sound frequencies). The gray shade corresponds to the attenuation magnitudes in dB. Range of $\varphi \in (0, 180)$ denotes the right half (ipsi-lateral) of auditory space, range of $\varphi \in (180, 360)$ covers the left (contra-lateral) half-plane.

vapid character of the sound in comparison to the other methods. Despite all the above-mentioned facts, the ratio of the complexity of the positioning algorithm and its final effectiveness is still sufficient in most cases.

The HRTF behavior is frequency dependent. All the spectral features such as peaks, notches (local minima and maxima in the magnitude of the transfer function) change its position according to varying source position. In order to demonstrate some of the dependencies, see Fig. 4.3. The heat map represents the magnitude of the right channel HRTF of various spectral components in a range of 360° in the horizontal plane analogically to the panning method in Fig. 4.2. Each vertical slice represents logarithmic spectrum of the HRTF corresponding to φ position of the sound source in the horizontal plane (around the listener, eye-level). Frequency dependence is obvious in contrast to the AP. While low frequency components up to approximately 1.5 - 2 kHz remain essentially constant, higher band above 5 kHz shows distinctive variation. Apparent is the attenuation for the contra-lateral position, where for $\varphi = 270°$ several damping areas occur particularly in higher frequencies above 8 kHz. This effect is caused by head attenuation or reflection for shorter wavelength.

Longer wavelengths (comparable to head size) bend around a head, thus corresponding band in the transfer function for the contra-lateral position still preserves greater magnitude. When the front ($\varphi = 0°$) and back ($\varphi = 180°$) HRTFs are compared, attenuation of higher frequencies above approximately 9 kHz is distinct. This position-dependent inter-channel coloration is needed for localization ability in the horizontal plane (front-back) [9]. As stated in Section 2.3 the shape of the HRTF is subject dependent [14].

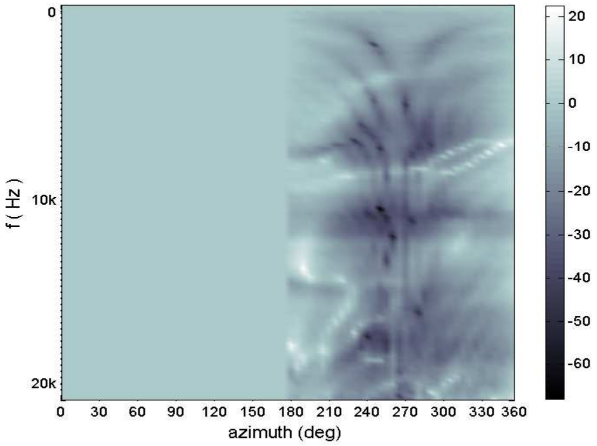

Figure 4.4: DHRTF heat map. This differential HRTF heat map introduces the character of particular spectral components within a 360° horizontal plane (step by 5°). Since this demonstration corresponds to the *right* ear, the transfer function remains constant for the ipsi-lateral position of the source. The gray-scale bar corresponds to the magnitude in dB.

The frequency-dependent character of the right channel transfer function of the DHRTF is demonstrated in Figure 4.4. It is important to emphasize that the essence of the DHRTF method is that the ipsi-lateral channel gain remains always constant, since all the processing occurs is only in the contra-lateral channel (see Eq. (3.11) and Eq. (3.12)). Therefore, the left half-space corresponds to constant 0 dB level. As with the HRTF, the most significant attenuation of magnitude (note in this case directly introducing the ILD) corresponds to source position around $\varphi = 270°$. Frequency areas around 7 kHz, 10 kHz and 17 kHz are attenuated at most. Very interesting feature of the contra-lateral side is an occurrence of the negative ILD at particular frequency areas, in Fig. 4.4 denoted by the lightest shades of gray color (e.g. at around 13 kHz at position $\varphi = 270°$).

By processing of the transfer functions of both channels for the three described methods, a position-dependent ILD is obtained for full 360 degree range in the horizontal plane. Fig. 4.5 introduces resulting ILD in the signal for AP (a), HRTF positioning method (b), and DHRTF positioning method (c). While panning method does not provide any frequency dependence, the dependence of the ILD within the DHRTF and HRTF methods on frequency is obvious. As expected, and as results from the attributes of the DHRTF, the features of panel (b) and (c) are completely identical despite each one corresponds to different method. The sharp discontinuity line in the center corresponds to the definition itself. When ears interchange their positions of 0° and 180°, the ILD features are reversed, thus the edge is more apparent. Moreover, the presence of the negative ILD for the positions close to the center (0°) emphasizes this effect. For the purpose of preserving the same scale in this graphical representation, the ILD of the AP has been adjusted, since the zero gain of the contra-lateral channel results in divergence to infinity in the limit case.

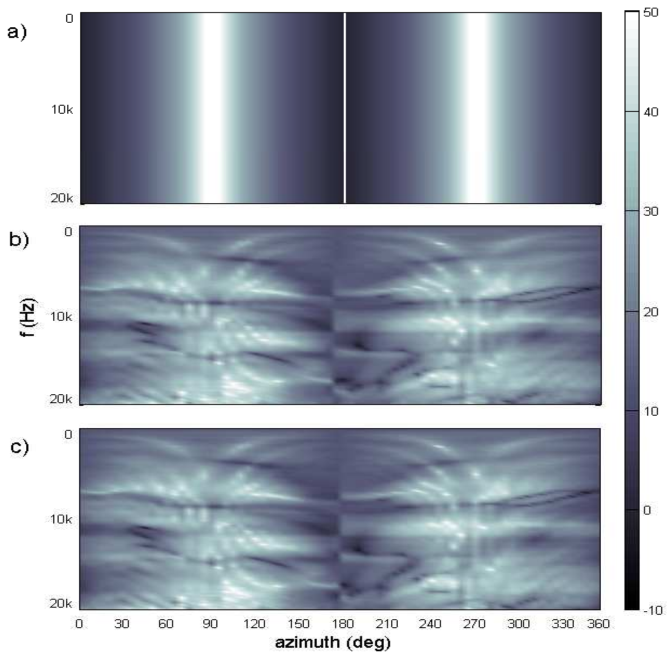

Figure 4.5: ILD azimuth functions. The gray map describes the frequency dependence of the ILD across positions of AP (a), HRTF (b), and Differential HRTF (c) for a 360° horizontal plane. The bar corresponds to the magnitude in dB.

4.1.1.3 Time Domain Aspects

To enhance the scope of analysis, the effect of each method is demonstrated on 10-period sine burst. This artificial signal was chosen for clarity of presentation and positioned by all the methods to a default position of $\varphi = 70°$. Resulting parameters of such stereo signal are shown in Fig. 4.6. Panning method provides different amplitude for each channel. Onset part of the sine wave starts in the same moment in both channels; the ITD is zero. In contrast, the HRTF and DHRTF methods deliver mutual time shift between the channels.

While the HRTF method provides distinct offset of the signal as a consequence of convolution in each channel, addition of the *convolution tail* in the DHRTF method affects only the processed channel. Note the tail is related to length of the dHRIR (actually HRIR) and is affected by the present ITD as well. This fact may introduce restriction to the maximal length of the dHRIR in signal processing. When convolution tail is *too long*, it may be perceptible at the end of the signal and distract source localization. In standard HRTF databases the length of the HRTF varies usually within an interval from 3 to 6 ms (e.g. 4.54 ms in used CIPIC database [17]). In the next step it is necessary to perform listening tests in order to probe the final perception effect of the above-mentioned spectral and time features. The details of the listening test are specified in the following section.

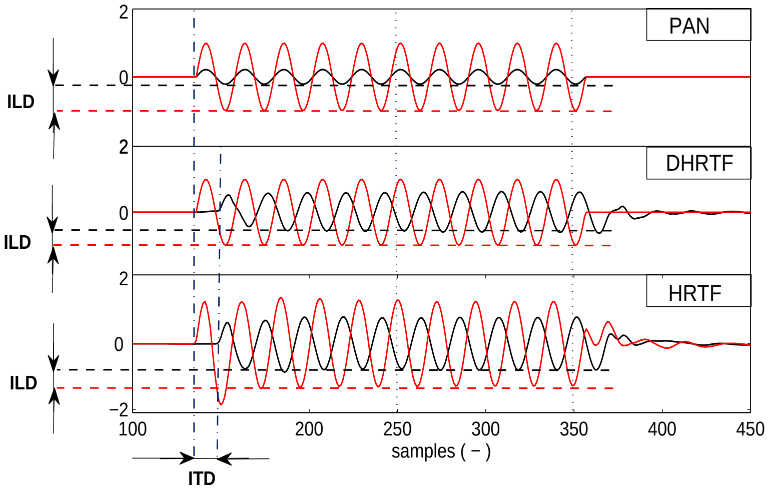

Figure 4.6: Positioning methods on sine wave. The effects of the presented methods are demonstrated on a 10-period sine burst. For a particular frequency, the DHRTF and the HRTF provide the same ILD and inter-channel time shift. ITD is absent in the AP. Note different signal offsets and onsets.

4.1.2 Subjective Comparison of Positioning Methods

In order to investigate how the stimuli positioned by the DHRTF method are perceived by the listeners and what is the difference in perception compared to the other two positioning methods, subjective listening tests were performed. The outputs of the objective comparison of the methods resulted in the selection of three parameters to be assessed in the listening test. The parameters were not primarily focused on the investigation of localization precision or the JND (Just Noticeable Difference), since the JND has been briefly investigated in previous work [A.6] with sufficient results. The factors to be assessed and rated by the subjects were as follows:

- **Spatial impression** represents the effect of spatial fidelity and credibility of the sound source located at particular positions; i.e. natural sounding.

- **Coloration** regards affection of the sound timbre. The main goal was to verify whether the DHRTF would incline to disturbing coloration of the final positioned sound due to the one-channel filtering.

- **Loudness** was expected to vary along particular positions according to Fig. 4.1. Varying loudness might be perceptible specifically when the positioned sound source moves.

4.1.2.1 Listening Test: Positioning Methods Comparison

A graphical user interface was designed and used for presenting stimuli to the subject and gathering the subjects' responses. Each trial of the test consisted of presenting four stimuli to the subject; three positioned stimuli to be assessed and one monaural reference stimulus (the original sound to be positioned). The reference was always presented first and the order of the following samples positioned by particular methods was randomized. After the initial presentation of all the stimuli the subject had unlimited option to listen to the presented sounds again by clicking on buttons corresponding to particular sounds. The subject was asked to adjust the value of *sliders* representing particular parameters (spatial impression, coloration, loudness) of each *unknown* positioning method. The slide scale consisted of 0.5 interval steps from 1 to 5 and were identical for all the three parameters. Verbal equivalents of slider value ratings are summarized in Table 1 in exact wording, as they were presented in subjects' instructions. This approach was chosen according to recommendations in [72]. Finally, the subject was asked to select the most preferred stimulus intuitively according to the quality of spatial impression and natural character of the sound.

4.1.2.2 Stimuli Description

Three different stimuli, with lengths ranging from 1.6 s to 3.4 s, were chosen for the test; snare drum phrase, speech segment, guitar chord. Each of the stimuli was positioned using the particular methods: for AP the samples of each channel were multiplied by corresponding gains, and direct convolution of the stimuli and 200-samples long filter response (FIR of order 199) was implemented for the HRTF and DHRTF. The convolution within the DHRTF was performed only for one channel, as results from its definition. Spatial division for the front half-plane was chosen simple in range from φ = -90° to φ = +90° with step of $\Delta\varphi$ = 30°. Therefore, the 3 stimuli and 7 positions result in 21 trials of the test. Each trial was expected to last no more than 1 minute, thus the session length did not exceed 25 minutes in order to maintain the subjects' motivation to fulfill the task correctly [72].

The length of the original HRTF data set as well as the resulting DHRTF data set consisted of 200 samples of standard sampling frequency 44.1 kHz. Therefore, the maximal time length of HRIR corresponds to 4.54 ms. The DHRTFs were selected from two available HRTF sets for acoustic manikin [17] in order to avoid the spectral spike occurrence. It is important to notice that in terms of assessing loudness, mutual gain of the particular methods was normalized to the same mean gain. The gains are shown in Fig. 4.1. It is also assumed that the differences within the methods are much more significant than differences resulting from occasional deviations of the subjects' anthropometric parameters from the manikin's [14]. Therefore, subjective dependences of the individual HRTFs were not taken into account.

SCORE	SPATIAL IMPRESSION		COLORATION		LOUDNESS	
	RATING	DESCRIPTION	RATING	DESCRIPTION	RATING	DESCRIPTION
1	very poor	dull sound inside the head	much worse	timbre is much different much worse than the reference	well perceptible	DECREASE of loudness
2	poor	-	slightly worse	-	barely perceptible	DECREASE of loudness
3	average	credible source position but no natural character	inaudible	the same timbre perception of both stimuli	same impression	NO CHANGE of loudness
4	good	-	slightly better	-	barely perceptible	INCREASE of loudness
5	excellent	sound outside the head in specific position	much better	timbre is different - much better than the reference	well perceptible	INCREASE of loudness

Table 4.1: Rating keys. Word expressions corresponding to numeric values were assigned to the assessed segments.

4.1.3 Results

The test was performed on 26 subjects, aged from 19 to 43. Both musically skilled subjects and people with no musical background were included in this set. The results were statistically analyzed by the software GraphPad Prism. The following graphs present the results of each assessed parameter by *boxes* representing 25% to 75% percentiles and *whiskers* showing the sample standard deviation. The mean value of each data set is represented by a horizontal line in the box. The results were subjected to multiple factor analysis of variance, RM-ANOVA (Repeated Measures Analysis of Variance), with two factors: *positioning method* (AP, HRTF, DHRTF) and *position* ($\varphi \in \{-90, -60, -30, -0, +30, +60, +90\}$).

Fig. 4.7 shows the results for the ratings of *spatial impression*. The most weak spatial effect was provided by the *amplitude panning* method, while the best results of spatial depth were produced by the HRTF method. The results of the DHRTF method appear in the middle range, inclining more to the character of the HRTF. The results of the spatial impression parameter formed a *V-shape*, with the tip of the ”V” letter pointing to $\varphi = 0°$ as the cues for perception of the space depth are connected to the synchronous ITD and ILD. This occurs more robust towards to the side positions. For the central position of $\varphi = 0°$ the average values are almost identical. The analysis of variance revealed the following statistical outcomes: for variance within the *method* $F(2,24) = 109.1$, $p < 0.0001$, and for variance within *position* $F(6,20) = 11.08$, $p < 0.0001$. This refers that both factors *positioning method* and *position* are statistically significant in rating of spatial impression.. Results for channel *coloration* are shown in Fig. 4.8. The line at the value of 3 denoted *Imp.* on the y-axis refers to the level of imperceptibility. Despite the fact that the AP is the only method, which does not include channel filtering, its rating is inferior to the other two methods, specifically from the side positions. This effect is probably connected with the unnatural character of the sound resulting from a close-to-zero gain in the contralateral channel in these positions. The HRTF and DHRTF have comparable values of their means along the azimuth. However, the deviation of the HRTF is remarkably higher.

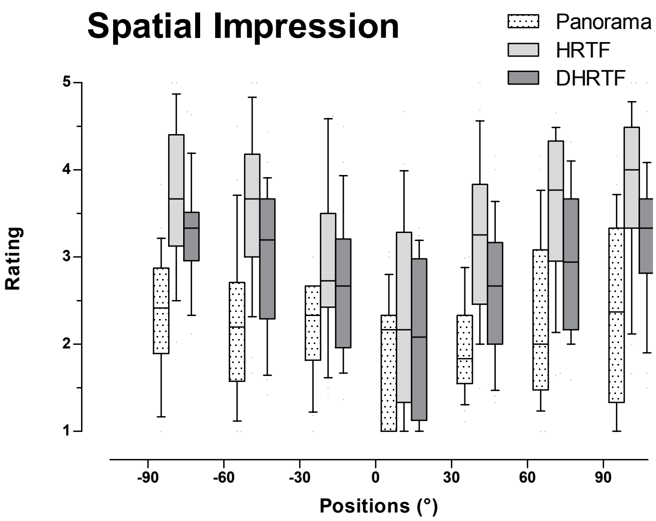

Figure 4.7: Spatial Impression rating. These results show ratings of the parameter *spatial impression*. The *amplitude panning* shows the lowest spatial effect beside the HRTF, which provides the best results. The DHRTF inclines more to the attributes of the HRTF, with a slightly worse impression. The *boxes* represent 25% to 75% percentiles and the *whiskers* show the standard deviation. The mean value of each data set is represented by a horizontal line in the box.

This phenomenon results from easily perceived stimuli timbre changes within the HRTF method that is caused by a boost in mid frequencies of the positioned sound (see the gain of the ipsi-lateral HRTF, right, in Fig. 3.12 (a)). This effect was assessed by both *better* and *worse* options, specifically, when musically skilled subjects preferred the mid-boost character. The DHRTF method preserves the original timbre of the stimuli the most against the previous hypothesis. This is most likely caused by maintaining the unprocessed channel as dominant resulting in the perception of the sound timbre close to the origin even in side positions, where the difference is maximal. RM-ANOVA revealed the following outcomes for variance within *method* $F(2{,}24) = 60.67$, $p < 0.0001$, and for variance within *position* $F(6{,}20) = 1.18$, $p = 0.32$. The values refer that only the factor of *positioning method* is statistically significant for *coloration* parameter. Slight trend of dependence on *position* is observable for AP, when the outer positions are assessed worse, probably due to the unnatural character.

Regarding *loudness* assessment, see the graph in Fig. 4.9. The perception of loudness did not vary significantly across the positions. In accordance with the total gain curves (see Fig. 4.1) a slight rise for the DHRTF and small decay for the HRTF at central position is noticeable. The gains for all the methods were aligned using the same mean value; however, the results for loudness variation show a difference. It is important to note that the total gains for the HRTF and DHRTF methods were derived based on the energy of their frequency-dependent impulse response. Under normal conditions the loudness perception depends also on the spectral character of the processed sound. A typical shape of the HRTF

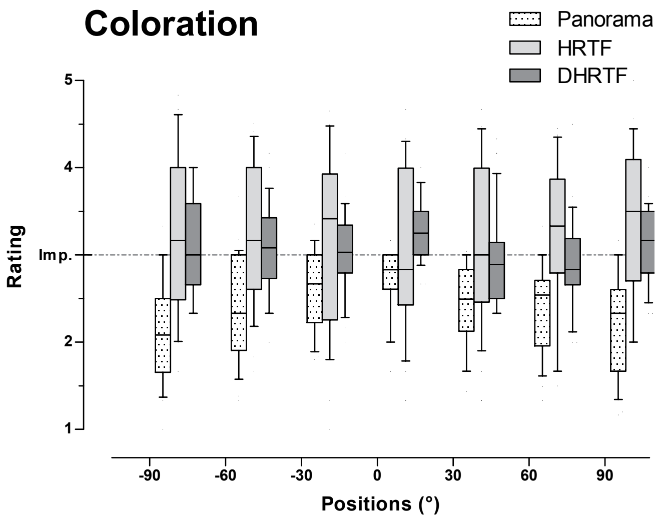

Figure 4.8: Coloration rating. Results are shown for the *coloration* parameter. Though the average values of DHRTF and HRTF are similar, coloration is more affected by the HRTF according to the larger variation of the rating. This timbre change is either preferred or rated as worse. The unnatural character also probably contributes to the low rating of the AP method. The *boxes* and *whiskers* are analogical to Fig. 4.7.

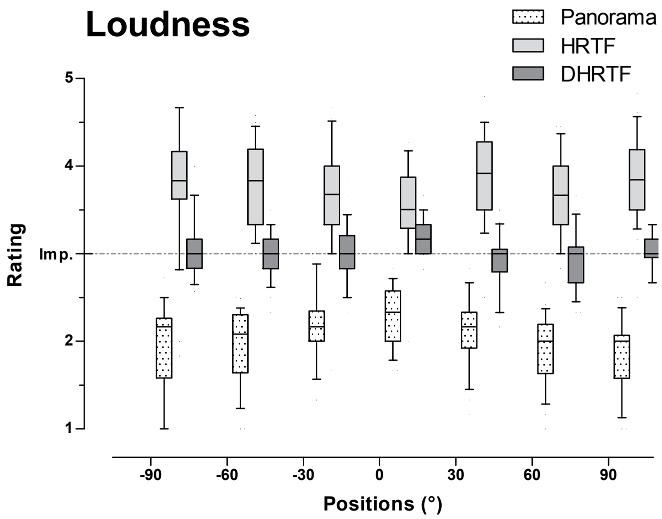

Figure 4.9: Loudness rating. Results are shown for the *loudness* parameter. A boost of middle frequencies causes primarily higher perception of loudness in the HRTF method. In the DHRTF method, the middle frequencies differed minimally. The *boxes* and *whiskers* are analogical to Fig. 4.7.

	METHOD **AP**	 **HRTF**	 **DHRTF**
Processing requirements	multiplication 2 channels	convolution 2 channels	convolution 1 channel
Spatial impression	poor	excellent	excellent
Various elevation	no	yes	no
Front/back positioning	no	yes	partly
Channel coloration	none	high	tiny

Table 4.2: Methods summary. Attributes of particular positioning methods are summarized here.

contains a resonance peak between 4 and 8 kHz (see Fig. 3.12 (a)), which corresponds to the most sensitive area of the human ear [7]. This results in the previously discussed mid-boost effect, which may be finally reflected as an increased perception of loudness. The decreased AP rating is probably also a result of spectral independence of the changes. The perception of loudness variance for the DHRTF method is minimal, except for the small increase in the central position. The analysis of variance revealed the following outputs: for variance within *method* $F(2{,}24) = 706.1$, $p < 0.0001$, and for variance within *position* $F(6{,}20) = 1.23$, value $p = 0.30$. This refers that only the factor of *positioning method* is statistically significant. The results disprove the previous hypothesis for the DHRTF that the loudness will fluctuate significantly along the positions due to the non-uniform (one-channel) filtering. Despite the examined positions were roughly distributed in the frontal plane, a follow-up experiments performed within [A.1] confirmed this statement by employing moving virtual sound objects.

The last task of each trial of the test was to select the most preferred stimuli. The results presented in Fig. 4.10 show that the method preferences were not consistent within stimuli and this is possibly related to their spectral content. For sharp stimuli with strong high-frequency content such as the snare drum phrase (a) the HRTF positioning resulted in a strong boost and an even more sharpened sound. Therefore, the *milder* DHRTF was mostly chosen in this case. However, the *high-band* and *mid-band* enhancement may have even improved the entire stimuli sounding, as in the case of the guitar chord sound (c) due to its tonal character. This effect contributed to a good spatial quality, thus the HRTF was selected by the majority in this case. The subjects preferred mostly the DHRTF also for the male speech fragment (b). In the final summary (d), the DHRTF and HRTF were most preferred and basically equal, compared to the AP method (DHRTF 46.3%, HRTF 46.5%, AP 7.2%), which was preferred only by a minority.

The attributes of each method are summarized in Tab 4.2. While the *amplitude panning* offers simple implementation at the cost of poor spatial impression, the HRTF demonstrates a complex approach with good spatial results. The DHRTF enables the reduction of processing to only one channel, while preserving remarkable spatial outputs and negligible channel coloration.

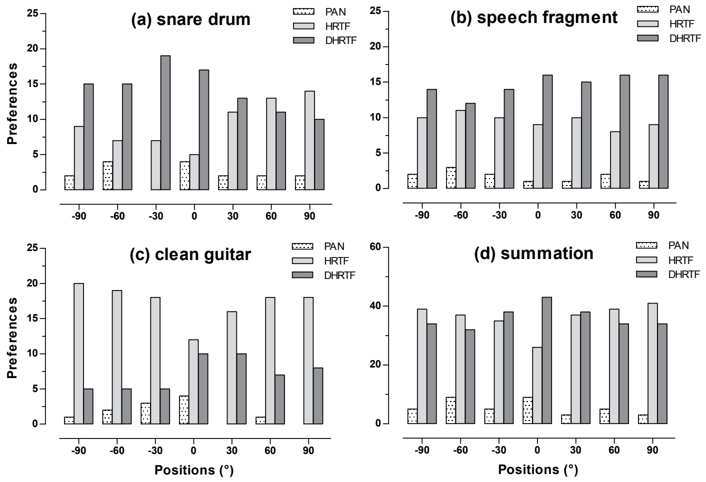

Figure 4.10: Preferences based on content. Histograms of the method preference that depended on the character of the stimuli are shown. Sharp sounds with high-frequency content are sensitive to significant boosting by the HRTF. This effect might be desirable for tonal instrument characters (i.e. guitar).

4.1.4 Discussion

The DHRTF based method can have useful applications in headphone listening. Any listener may expect sound reproduction to have the following qualities: it is pleasant, it feels natural and it achieves the desired sound location perception. To test, how the DHRTF method satisfies these requirements, parameters related to the qualities described above were chosen: spatial impression relates to location effects; coloration captures both how pleasant and natural the sound is, albeit mostly for a trained ear; and loudness should change smoothly and in a sense that it is related to all the qualities mentioned above.

The objective analysis highlights points, where artifacts and noises can distort listening. The DHRTF method performed remarkably well in the subjective evaluation.

The DHRTF might prove advantageous in comparison with the HRTF. Two-channel processing may increase the requirements for computational resources, when multiple sources are simultaneously rendered. This situation might arise in computer games or in training assistive programs for the visually impaired [4, 60]. The DHRTF can be also effectively used in music post-processing (mixing), since the method provides very low timbre affection along with solid spatialization. Some other sound examples to test with the three methods can be found in the collection made available by R. O. Duda [10].

4.2 Implementation of Moving Source

This section focuses on aspects of implementation of moving virtual sound source in the virtual auditory environment, which was briefly introduced in 2.4.2. Rendering of dynamic source based on use of the HRTF or DHRTF is achieved by application of filtering with varying impulse response of the filter. The algorithm may be in its essence implemented by several methods, where each approach provides slightly different output signal resulting in variance of perception of spatial fidelity. The main difference results from different output of linear and circular convolution usually implemented in such processing. Particular variants are described, discussed and compared by subjective listening test of the positioned signals. Other aspects of implementation as artifact occurrence are also discussed.

When an application utilizing virtually positioned sound requires movement of the source [88, 57, 54, 83], it cannot be reached by a simple implementation of equation Eq. (2.11). Since this algorithm is based on summation of delayed copies weighed by appropriate coefficient of HRIR, it refers that the signal passes through this constant coefficient by its whole length. Therefore, filtering with variable coefficients of FIR filter response has to be implemented. A scheme of such filter is shown in Fig. 2.14; the architecture is similar to standard FIR filter except the coefficients vary in discrete time τ. The expression of time-varying filtering is presented by following equation

$$y_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n] = hrir_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n] * x_m[n] = \sum_{k=0}^{M-1} x_m[k] \cdot hrir_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n-k], \tag{4.4}$$

where angles φ and ϑ are function of discrete time τ. In order to preserve real-time varying ITD and ILD, it is necessary to process the signal by the varying impulse response. Virtual positioning is probably one of the most sensitive processes among processing tasks requiring precise synchronization of both channels. Specific features may occur during the processing in dependence on particular method. The following subsections introduce two processing algorithms and one combination technique for virtual positioning.

4.2.1 Dynamic Positioning with Linear Convolution

The first method involves segmentation of the original signal into segments of length of L samples and linear convolution of each part with appropriate HRIR of length M, as shown on the scheme in Fig. 4.11. This process can be described by Eq. (4.5), which assumes decomposition of the signal into Q elements

$$y_{\xi,lin}^{\varphi[\tau],\vartheta[\tau]}[n] = \sum_{\substack{q=0 \\ 0<n-qL<L}}^{Q} x_m[n-qL] * hrir_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n], \tag{4.5}$$

where q refers to the order of segment and L represents the length of the segment. Convolution of the segments and the responses results in a creation of new segments of length

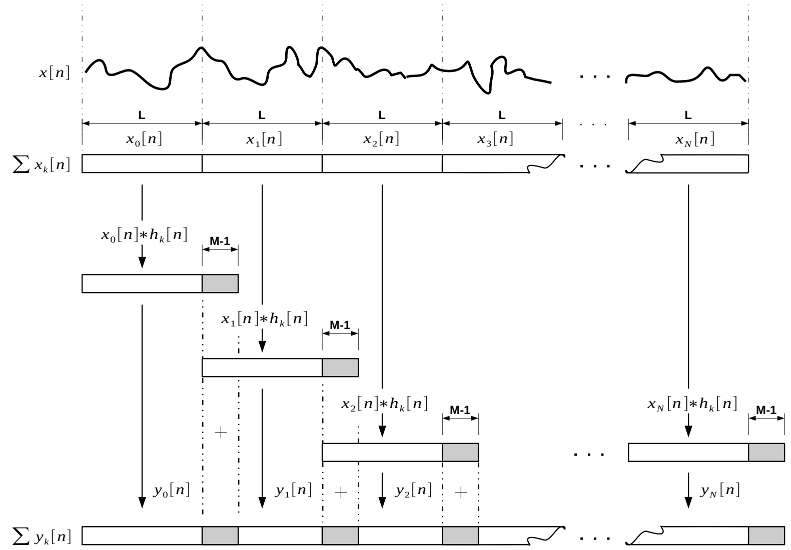

Figure 4.11: Linear convolution-based positioning. A scheme of block-by-block processing of linear convolution is demonstrated. Convolution *tail* of N-th segment is added to the following (N+1)th segment.

L+M-1 each [89]. Therefore, the final signal composition adds the extra M+1 long convolution tail of N-th segment to the onset of (N+1)th segment. This part of processing enables a smooth conjunction of all elements [48].

4.2.2 Dynamic Positioning with Circular Convolution

Another approach how to implement block processing is mutual spectral multiplication of the segments an appropriate complex transfer function. This algorithm is usually implemented as a standard in various signal processing applications [89, 48, 56, 55]. Operations with segments can be described by

$$\begin{aligned} y_{\xi,cir}^{\varphi[\tau],\vartheta[\tau]}[n] &= \sum_{\substack{q=0\\ 0<n-qL<L}}^{Q} x_m[n-qL] \otimes hrir_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n] \\ &= \sum_{\substack{q=0\\ 0<n-qL<L}}^{Q} \mathrm{FT}^{-1}\Big\{\mathrm{FT}\{x_m[n-qL]\}\cdot\mathrm{FT}\{hrir_{\xi}^{\varphi[\tau],\vartheta[\tau]}[n]\}\Big\}. \end{aligned} \qquad (4.6)$$

The method of spectral multiplication strictly requires the same length of the impulse response and the signal segment. When this requirement is not fulfilled, appending of zeros to the signal is an option [48]. The length of the HRIR inclines usually to 150-300

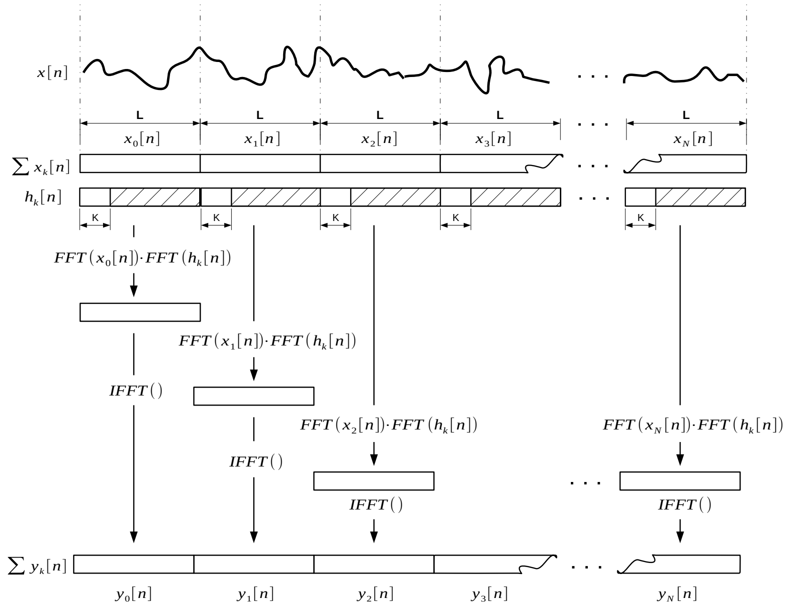

Figure 4.12: Circular convolution-based positioning. A scheme of block-by-block processing of circular convolution implemented by FFT is demonstrated.

samples in standard $f_s = 44.1$ kHz [61, 17]. Due to non-negligible artifacts, which may occur while short window length is applied in the processing (e.g. 128 samples), it is assumed that the impulse response (i.e. HRIR) of length K is always shorter than signal segment of length L cut from the origin. This fact offers simple addition of zeros to the impulse response in order to extend the length of the segment. The value of segment size is usually chosen from set of power of two due to reducing computing power consumption [89]. A scheme of this algorithm is introduced in Fig. 4.12. Each processing method results in slightly different outcome regarding particular onset and offset of the segment signal. In comparison to previous method see Fig. 4.13, where the procedure is implemented by both methods to the test signal (2-channel, 10-period sine wave, $f = 440$ Hz, $f_s = 44.1$ kHz). Although main features regarding amplitude ratio and phase shift are preserved, onset and offset differ.

4.2.3 Employing Overlap-Add Method

Method of smooth assembling of the processed segments is well known as OLA/OA (overlap-add) technique [89]. The method is based on weighting of each segment by window function $w[n]$. While direct assembling of the processed segments in particular for method described in section 4.2.2 usually results in perceptible artifacts, OLA/OA enables smooth assembling resulting from the weighting process. This process can be applied to both methods. The following equation describes its application to circular convolution

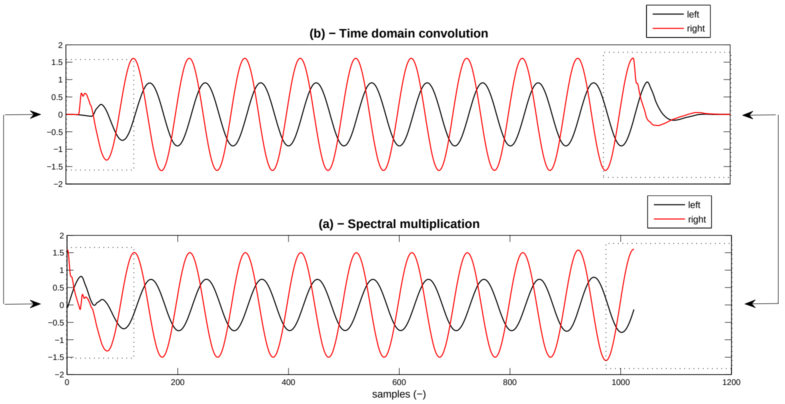

Figure 4.13: One-block demonstration of the processing. Different onsets and offsets of the signal processed by circular and linear convolutions are depicted. This phenomenon is negligible in static processing, however essential in dynamic processing.

implemented by FFT

$$y^{\varphi[\tau],\vartheta[\tau]}_{\xi,lin}[n] = \sum_{\substack{q=0\\0<n-\frac{qL}{2}<L}}^{Q_o} \left(\left(x_m[n-\frac{qL}{2}]*hrir_\xi^{\varphi[\tau],\vartheta[\tau]}[n]\right) \cdot w[k-\frac{qL}{2}]\right), \tag{4.7}$$

where $w[n]$ stands for weighting function of shifting window. The equation assumes overlapping of half of the window length, i.e. L/2. However, the overlapping factor may vary from 1 to N-1 samples. The algorithm is graphically illustrated in Fig. 4.14. There are many variants of the weighing window, for our purposes standard Hanning window was used. Hanning window can be defined as

$$w_h[n] = \sin^2\left(\frac{\pi n}{N-1}\right). \tag{4.8}$$

4.2.4 Effect of Window Length

Previous sections introduced processing methods handling segments of length L. Setting the window length is also an important parameter for the processing itself and attributes of the moving source. Specific boundaries appear for the maximal and minimal sizes. The lower boundary of window size L is determined by subjective quality of processing regarding artifacts occurrence. Since small size of window ($\approx$ 256 samples) results in well perceptible artifacts of particular implementation, an effort for setting window length as large as possible may seem reasonable. However, window size is also limited from upper boundary by maximum angular frequency (projection of the movement to horizontal or median plane), which the positioned source is enabled to reach. Otherwise, the final

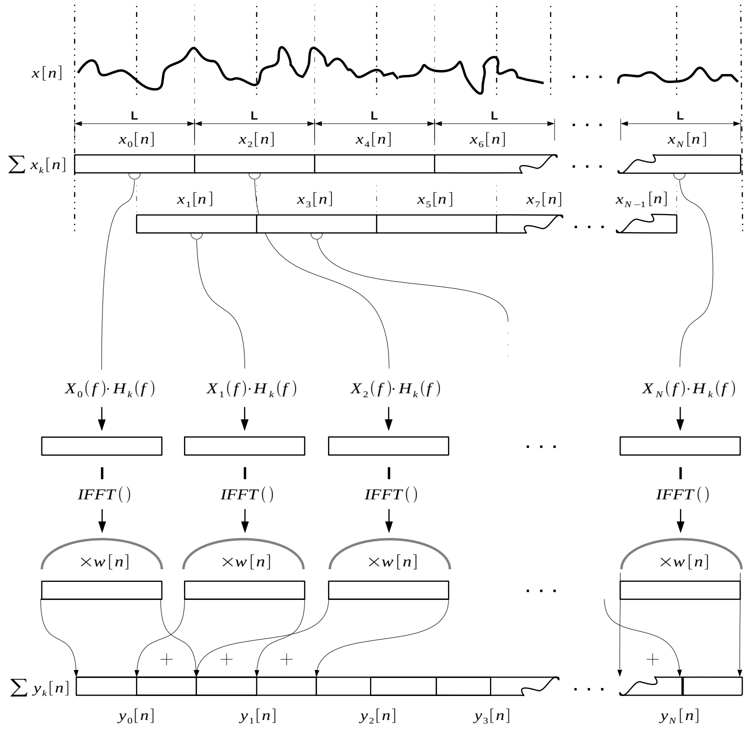

Figure 4.14: Overlapping of segments. The overlap-add (OLA/OA) method applied to virtual positioning is implemented by segmentation processing of circular convolution. The overlapping factor is 0.5.

perception of continuous natural shift of the source position is distorted by occurrence of discrete position jitters. In order to avoid this phenomenon, it is necessary to take these aspects into account while implementing virtual positioning of the moving source. For more comprehensive scope assume geometric arrangement as in Fig. 4.15. Real measured HRTF is usually sampled by discrete spatial locations of angular step $\Delta\vartheta$ (another approach e.g. [90]). For the explanation purposes, the situation is now simplified and reduced only to azimuth plane, though virtual positioning concerns both horizontal and vertical movements. For spatial fidelity of moving source and continuous switching of HRIRs it is necessary to ensure that time length of one segment Δt is shorter than time of movement of the source about angular element $\Delta\vartheta$, as states here

$$\omega_r \leqslant \frac{\Delta\varphi}{\Delta t} = \frac{\varphi_n - \varphi_{n-1}}{T_s \cdot L \cdot (1-o)} \tag{4.9}$$

where ω_r stands for radial frequency, T_s represents reversed value of sampling frequency, L is length of the segment and o represents overlap factor (in this study implicit $o = 0.5$). Therefore, after fulfilling this requirement, each HRIR is utilized to process at least one

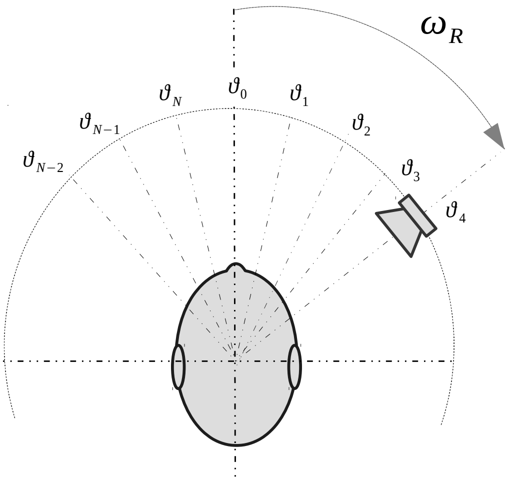

Figure 4.15: Discrete HRTF positions. Switching particular discrete positions of $hrir_{\xi}^{\varphi_k,\vartheta_k}$ for φ_k and ϑ_k according to the sound source movement is shown.

segment. If not, position of the source may abruptly turn from the initial position to farther one. This phenomenon results in perception of disturbing jitter of the source position. Validity of this equation was verified by performance of listening test focused on distinctiveness of position leaps in the signal, as shown in the following section. It is worth to notice that for real application, the edge angular frequency does not concern only literally speed of the sound source itself (e.g. task of positioning of flying-around object in action computer game or artistic impression). A demonstration of a *bank* of impulse responses for switching the HRIRs is demonstrated in Fig. 4.16.

Furthermore, virtual reality often requires connection with head-tracking system [43], [91]. Purpose of such system is to detect head (or body-and-head) movement due to its compensation and consequent placing of the source to *absolute*, not only *relative* virtual position. Therefore, maximum angular frequency allowing continuous position change determines also maximum angular frequency of head turn conditioned by proper performance of the system.

4.2.5 Analysis of Dynamic Positioning

As mentioned in the first section, virtual positioning is a specific signal processing task requiring a high-precision approach of implementation. This section introduces results that have been discovered during investigation of the influence of various aspects of implementation. Both technical analysis and subjective assessment by listening tests were performed. Each of the processing method provides slightly different signal from technical point of

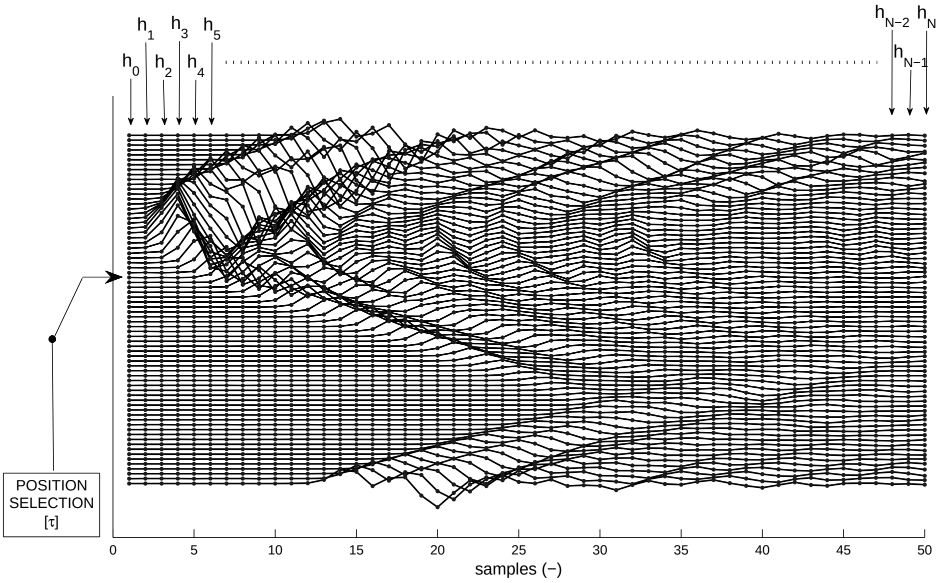

Figure 4.16: Switching FIR impulse responses. A bank of $hrir_{\xi}^{\varphi}$ for 360-radius spread as the variable response of a FIR filter is visualized.

view. Therefore, each segment of the signal contains modified information partly resulting form different method, partly from occurring artifacts. The essential task is to find whether such difference may result in a different spatial impression. Due to very complex behavior of HHS regarding spatial localization of sound source, it is important to compare objective results and subjective perception results. Comparison of different processing settings is aimed at the following criteria:

- Dependence of implementation on final spatial impression, i.e. width of the source position and extent of externalization of the sound.
- Audibility of processing artifacts, such as clicks or noises, which are concerning only purity of the signal with no relation to the spatial perception.
- Distinctiveness of position jittering related to the ratio of angular frequency and length of the processing segment.

The second and third points are technically very uneasy to detect (implementation does not outweigh results), thus these aspects are restricted only to listening tests.

Informal subjective tests and comparisons of the signals referred that processing methods based on circular and linear convolution provide sound of the same character regarding timbre perception that expected by means of filtering process. When only mono signals of left or right channel were compared within the methods, no difference was perceptible. However, the final spatial perception seemed to be different within space depth for both

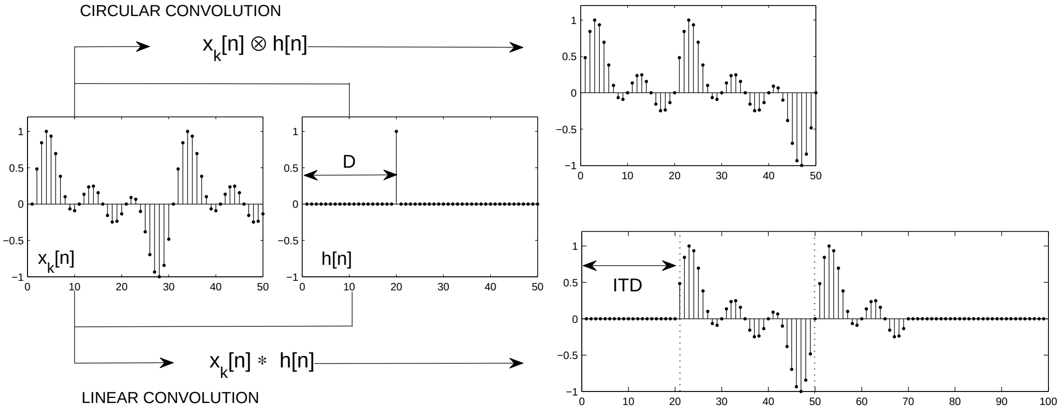

Figure 4.17: Convolution of a segment. A demonstration of the resulting effects of circular (top) and linear (bottom) convolutions on the input signal for the purposes of virtual positioning is depicted.

circular and linear convolution, when both channels were presented as a stereo file. Although spectral envelope of the whole signal was identical, time domain waveform showed a specific time shift in each segment. The time shift results in occurrence of differential signal, when the two signals were subtracted by each other. Note varying gain of each channel corresponds to varying source position. Since the HHS is minority-sensitive to phase of a complex signal [88], the method-based difference is not perceptible when only a mono signal is processed. However, this difference matters in terms of binaural processing of virtual sound source. Further analysis based on application of cross-correlations and differential spectrograms discovered the essence of this phenomenon. Figure 4.17 introduces the final effect of each method on a signal segment. In this example impulse response of the filter is defined as time shifted discrete delta function; therefore, $h[n] = \delta[n - D]$. While the linear convolution provides actual ITD of the signal, the circular convolution results only in phase shift of the in-segment signal (i.e. its tail part appears as an onset of the processed signal). As mentioned in the first section, ITD is one of the key-factors for estimation of source position in azimuth plane. The linear and circular convolutions may also provide the same output under certain conditions [89], thus this phenomenon is usually not necessary to be taken in considerations in static virtual positioning due to sufficient zero-offset.

4.2.6 Listening Tests: Dynamic Positioning

In order to verify the above-mentioned statements resulting form signal analysis and preliminary auditory impressions, series of listening tests were arranged. Virtual positioning process was applied to several real signal elements (speech, snare drum roll, saxophone phrase, vocal line, and clean rhythm guitar phrase) and one artificial wide-band signal occasionally recommended for localization tests [92] (i.e. white Gaussian noise modulated by sine wave of frequency 20 Hz). Sample frequency was chosen standard 44.1 kHz with 200-samples length of HRIR ($\approx 4.5\mu s$) of acoustic manikin from database available online [17]. Each signal was positioned in order to make one virtual turn-around from 0° to 360° around the listener. Angular frequency of the segments was chosen in values of 1, 2, and 4 $\mathrm{rad} \cdot \mathrm{s}^{-1}$. Length of the segment window L varied within a set of power of two from 256 to 16 384 samples. The test was finally performed on 11 subjects, both on those who already have encountered virtually positioned sound and those who have not encountered it yet. In the first part of the listening test, each subject was asked upon a method preference by assessing two versions of positioned stimuli of the same character. The task was to determine which stimuli shows better width, depth, and nature of spatial impression. Available options were chosen according to modified paired comparison method, i.e. the subject was also allowed to select *indistinguishable* [72]. The results are presented in Table 4.3 and Fig. 4.19.

4.2.7 Results

Segment processing by the linear convolution provides conclusively better spatial effects, although some subjects were not able to distinguish the difference. This had a strong impact on character of inter-subject responses; while the majority of subjects strictly assessed linear convolution better in all cases, several subjects preferred the methods randomly or were not able to distinguish among them. The second part of the test focused on distinctiveness of processing artifacts (i.e. usually noise-like or click-like) occurring due to insufficient window length and position leaps resulting from excessive window length. In this part the subject was asked to determine boundary window size represented by a point on scale in range from 1 to 7, where the artifact appears to be perceptible. The results are summarized in the graph in Fig. 4.18. Finally, the artifact curves consist of only results for overlapping-included methods, since the artifacts were perceptible constantly with use of direct conjunction of the segments both for the linear convolution and in the circular convolution method. Artifacts inclined to be well perceptible in tone-based stimuli (saxophone, vocal, guitar) and imperceptible in signals with strong transient (drums). Using the overlapping technique allows partial reduction of position jitters. The results show slight occurrence of the noise artifacts at length of 1024 samples and slight occurrence of position jitters at 4096 samples. Therefore, the recommended option of window length belongs to this interval with the best results for 2048 samples. Restriction concerning angular fre-

Option	**Preference (%)**
Circular convolution	9.2
Linear convolution	76.7
Indistinguishable	14.1

Table 4.3: Results of dynamic listening tests. The preference of positioning algorithms is summarized.

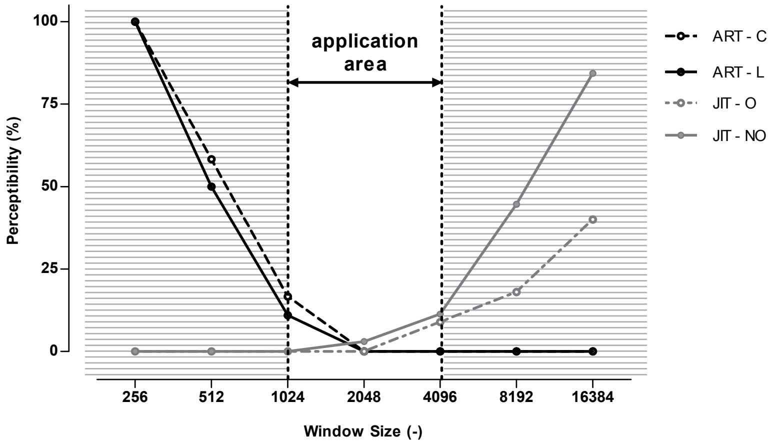

Figure 4.18: Application area. Dependence of perceptibility of processing artifacts (ART) and positioning jitter (JIT) on the size of segment window $w[n]$ is summarized here. (C-Circular, L-Linear, O-Overlapped, NO Non-overlapped).

quency confirms statement of Eq. (4.9). Continuous signals with no pauses included (e.g. modulated noise) incline more to be distorted by position jittering.

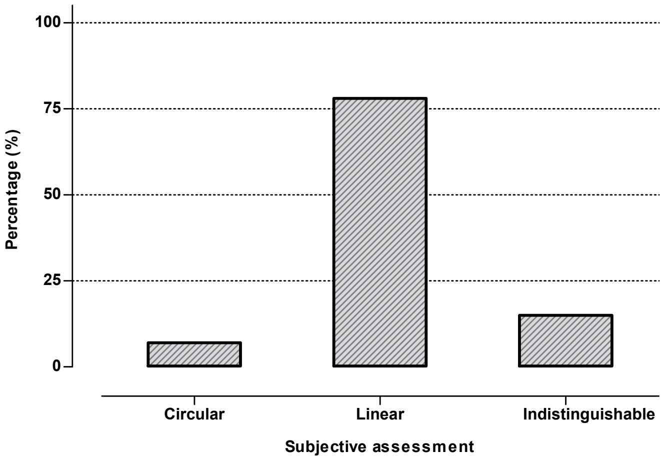

Figure 4.19: Circular vs linear convolution preference. Percentage of subject's preference of particular method is shown. Linear convolution was assessed with the highest preference rate.

4.2.8 Discussion

A more complex insight to the field of virtual sound source dynamic positioning has been introduced in this section. The main inspiration of this research results from a lack of detailed information of implementation in relevant publications regarding use of virtual positioning and a lack of information about this special narrowly-focused issue in well-known digital processing theory books or surveys. For static virtual positioning with fixed sound source, both implementation methods provide results with the similar spatial impression, since the essential channel onset may be affected only once. Furthermore, in this case, the output signals may be equivalent under particular conditions, when both impulse response and input signal include at least $M - 1$ zeros in the end [89].

Implementation of circular convolution may also be comfortable and sufficient for phase-independent or channel-independent processing (e.g. equalization in mixing console), where usually only spectral envelope of transfer function matters. However, for proper implementation of moving virtual source created by a set of the HRTFs, linear convolution is necessary to be implemented. If this requirement is not fulfilled, the occurring inter-channel differences, normally imperceptible under mono-listening conditions, will weaken the final spatial impression for most of the cases. This is primarily caused by absence of proper ITD, a crucial factor for localization cues. According to artifact occurrence, it is also recommended to involve overlap-add technique in the implementation algorithm. This work used overlapping factor $o = 0.5$ sufficiently. Length of L segment is recommended in range from 1024 to 4096 samples, ideally 2048. This step will help avoiding remaining processing artifacts and allow sufficient angular frequency of the source without significant leaps in source position. The algorithm efficiency can be furthermore enhanced by implementation of Doppler effect resulting in more natural effect of the sound movement as introduced in [93].

4.3 Perception Artifacts in the DHRTF Method

Previous effort of the author was focused on development of a positioning algorithm based on the Differential HRTF (DHRTF), which reduces processing requirements to only one channel instead of usual two for the HRTF positioning. According to the results published by the author in [A.2] (Sec. 4.1), the DHRTF method is capable of sufficient delivery of spatial impression and preservation of the natural character of the positioned sound. However, former experiments revealed that the DHRTF may cause artifacts in the positioned sound under specific circumstances (see Section 3.5). These artifacts primarily distort the timbre, but spatial perception is slightly affected as well. The goal of this experiment was to find a suitable easy-to-implement method of artifact reduction for the DHRTF algorithm and verify its efficiency. This section introduces three methods with different implementation requirements designed to reduce the artifacts occurrence. Their effectiveness is compared by means of subjective and objective tests. The following subsections describe the methods for artifact reduction, the setup of the listening tests and their performances. Finally, the results are presented.

4.3.1 Negative ILD and its Impact to DHRTF Positioning

The general assumption that the closer (ipsi-lateral) ear has always greater gain over the entire transfer function (i.e. the ILD is always positive) appears to not to be fulfilled in all cases. Negative ILD occurs in several specific cases leading to timbre and positioning artifacts. The cause of appearance of the negative ILD results from a unique relation of the two HRTFs in particular pairs. As briefly noticed in Sec. 3.5, the gain in the ipsi-lateral channel may be greater than in the contra-lateral on particular frequencies with a character of a narrow-band magnitude notch. For demonstration of such HRTF pair, see Fig. 3.12 (a). From the definition of the DHRTF, the same ILD is also present in its spectral module. Therefore, such constellation results in a sharp peak (spike) in the magnitude of the DHRTF exceeding the level of 0 dB, as shown in Fig. 3.12 (b). Since the essence of the DHRTF positioning lies in frequency-dependent attenuation and time delay of the contra-lateral channel, positive value of the logarithmic DHRTF magnitude results in *boosting* of the corresponding band in the contra-lateral channel. Notice that the perceived position of the stimuli always appears in the ipsi-lateral channel. Therefore, the effect of the mentioned DHRTF spectral spike leads into easily noticeable pure-tone character disturbance in the channel opposite to that in which the original source is perceived. However, the presence of the DHRTF spike (i.e. the negative ILD) is neither predetermined for specific spectral bands, nor for specific positions. Generally, the highest occurrence can be found around the central positions on the front-back axis (i.e. $\varphi = \pm 180^\circ$ and $\varphi = 0^\circ$). It also significantly varies among HRTF sets of particular subjects. Figure 3.12 (b) shows the DHRTF for the same position ($\varphi = 70^\circ$) obtained from the HRTF sets of two different subjects; the spectral spike occurs only in one of them.

For further investigation of the negative ILD occurrence, an analysis was performed over the whole database of the HRTF (43 subjects) available in [17]. The results are

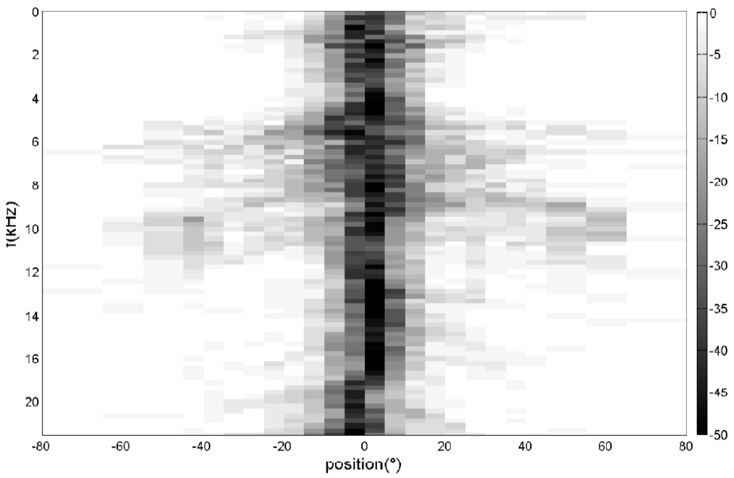

Figure 4.20: Histogram NILD. Histogram of the negative ILD occurrence for the horizontal plane measured over the HRTF sets from [17] is visualized. The gray scale represents the percentage of the negative ILD occurrence within the measured sets, N = 43.

demonstrated by histogram in Fig. 4.20. Not every DHRTF set is conditioned to be affected by the negative ILD. While the negative ILD around the central position ($\varphi \in \{0°, 180°\}$) occurs for specific frequency bands in up to 50 % of the sets, the amplitude of occurrence falls for the farther positions. For demonstration of spectral variety of the negative ILD, see Fig. 4.21, which shows specific positions of the negative ILD in several randomly selected sets. In order to avoid the artifacts in the positioned sound, either a set of appropriate HRTF can be selected, which may not be advisable due to anthropometric differences [14], or the artifacts can be reduced by a *suitable* method. Our approach to how to eliminate the artifacts is described in the following section.

4.3.2 Methods of Artifact Reduction

Since the DHRTF-based positioning method is not well-explored yet, it is necessary to develop and test methods for artifacts reduction. As mentioned above, the artifacts are tone-like elements occurring in the contra-lateral channel of the positioned sound. The artifacts are well perceptible within *static positioning*; when the sound source is fixed in particular position, the tone is well distinguishable. The situation is slightly different for *dynamic positioning*, when position of the virtual source changes. In this case, the signal is processed block-by-block [89], where each segment is convolved with appropriate dHRIR. Since the negative ILD mostly occurs at various positions for particular frequency bands (see Fig. 4.21) with its dominant part above 5 kHz, mostly higher frequencies are involved. Therefore, the final sound effect of the artifacts resulting from quick switching among

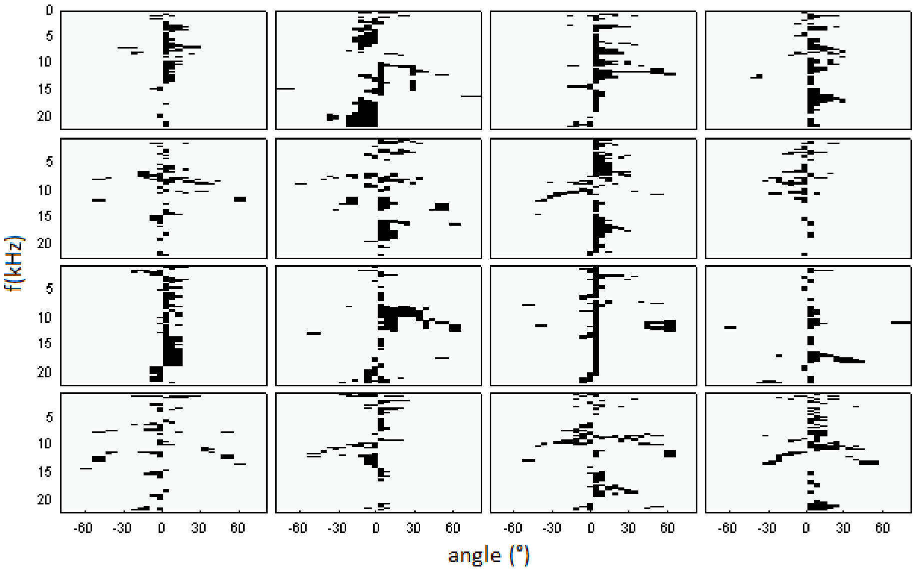

Figure 4.21: Artifacts occurrence. Several randomly chosen DHRTF sets with positive artifact occurrences created with the data from the database in [17] are visualized. The pattern of the negative ILD is unique for each subject as is the corresponding HRTF set. The negative ILD occurs mostly near the *front* and *back* positions ($\varphi = 0°$ and $\varphi = 180°$), where the inter-channel gain difference is weak. For better resolution only one half-plane is shown.

particular dHRIRs resembles the *birdies artifact* known from low-bit rate of perceptual (lossy) coding. From the principle of the DHRTF method, the artifacts occur always in the contra-lateral channel; therefore, in the other half-plane than the virtual source is located. Despite the artifact actually results from the positioned signal, it is perceived separately from the desired. The proposed methods of artifact reduction are supposed to eliminate the artifacts while preserving the original spatial arrangement and the original timbre of the source at the same time.

Independently on the selected method, the processing of the DHRTF is performed only within its magnitude, since the artifacts result primarily from the negative ILD. The reducing algorithms are based on extraction of the phase from the original $H_D^{\varphi}[\Omega]$ at the first stage, consequent processing of the module by particular algorithm(s), and finally introducing the original phase back to the signal. In the last step, inverse Fourier transform is applied to the DHRTF in order to get the dHRIR for implementation of the processing algorithm. This procedure is demonstrated in Fig. 4.24. More details concerning the issue of virtual positioning in the time or the spectral domain can be found in Section 4.2.

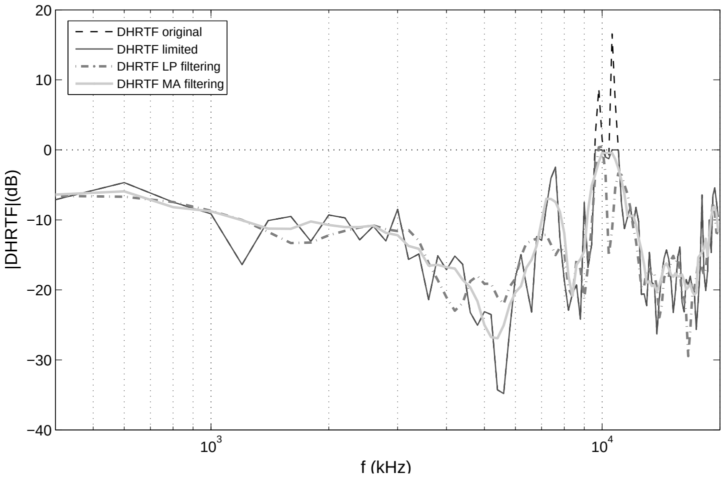

Figure 4.22: Elimination of artifacts. The impact of the proposed methods on a single DHRTF magnitude is demonstrated. *Smoothing* by the moving average follows the original DHRTF curve the best.

4.3.2.1 Spectral Amplitude Limiting

The most trivial approach is to remove all the spectral content in the DHRTF magnitude exceeding the reference threshold level. This can be achieved by performing hard limitation of the magnitude curve, where every spectral sample exceeding the threshold (i.e. the negative ILD spike) is reduced to this reference level. Such algorithm can be written as

$$\left|\widehat{H}_D^{\varphi}[\Omega]\right| = \begin{cases} \varepsilon & \text{for } \left|H_D^{\varphi}[\Omega]\right| > \varepsilon \\ \left|H_D^{\varphi}[\Omega]\right| & \text{elsewhere,} \end{cases} \tag{4.10}$$

where Ω denotes discrete frequency of the signal, $\widehat{H}_D^{\varphi}[\Omega]$ represents adjusted DHRTF for particular position φ, $H_D^{\varphi}[\Omega]$ represents initial DHRTF, and ε stands for the threshold level. The reference level was considered 0 dB in the logarithmic scale, which is equivalent to 1 in the linear scale.

Preliminary tests showed that the simple *spectral amplitude limiting* may sometimes appear insufficient due to remaining local maximums in the transfer function *under* the reference level ε. Therefore, the algorithm shall be extended by smoothening of the DHRTF magnitude while preserving ILD localization cues.

4.3.2.2 Low-pass Filtering of Spectral Amplitude

The next method employs multiplication of the spectrum of the DHRTF magnitude by weighing mask, eliminating the higher frequencies (i.e. sharp spectral shapes), which can be expressed as

$$\left|\widetilde{H}_D^{\varphi}[\Omega]\right| = \mathrm{FT}^{-1}\Big\{\mathrm{FT}\big\{\widehat{H}_D^{\varphi}[\Omega]\big\} \cdot \Pi[\xi]\Big\}, \tag{4.11}$$

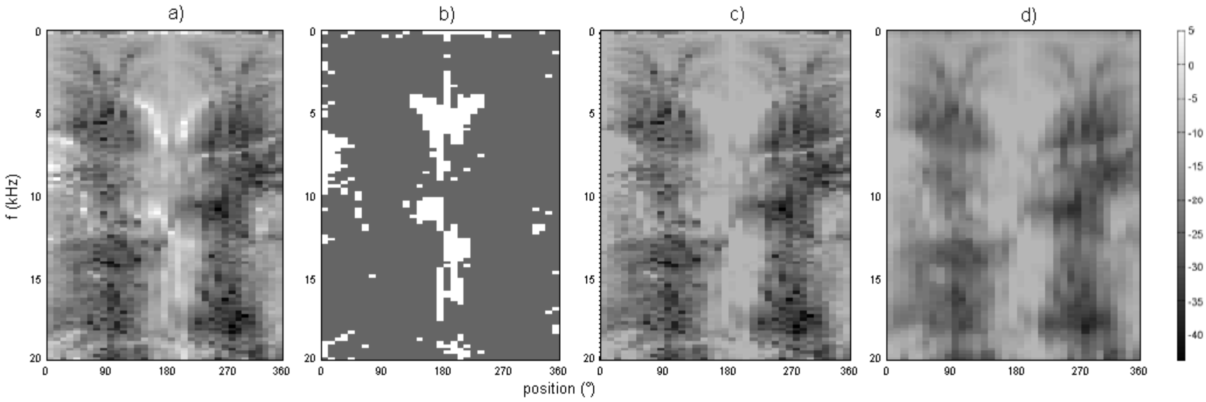

Figure 4.23: DHRTF adjustment. Graphic demonstration of particular steps in the DHRTF adjustment is presented. Original DHRTF (a), locations with negative ILD occurrence (b), hard limiting for $H_d[\Omega] > 0$ dB (c), filtering of the limited DHRTF (d). Since the output of *MA smoothing* and *low-pass filtering* is very similar in the resolution used here, only one is shown. Image softening in (d) corresponds to the smoothing of the local maximums of particular transfer functions.

where $\widetilde{H}_D^{\varphi}[\Omega]$ represents the limited and filtered DHRTF and $\Pi[\xi]$ denotes the weighing spectral mask for manipulation of the DHRTF spectrum. Parameter ξ represent a pseudo-frequency. The spectral mask $\Pi[\xi]$ is defined as a rectangular transfer function, as stated below (considered for linear scale).

$$\Pi[\xi] = \begin{cases} 1 & \text{for } \xi < \xi_0 \\ 0 & \text{elsewhere.} \end{cases} \tag{4.12}$$

The parameter ξ_0 represents a cut-off pseudo-frequency. This value is crucial for the final shape of adjusted DHRTF, as it determines the range of the low-pass filter. This method showed promising results during the preliminary tests. However, the high frequency cut-off may cause a new re-crossing of the DHRTF transfer function over the reference level ε.

4.3.2.3 Moving Average Smoothing of Spectral Amplitude

Due to the referred re-crossing disadvantage of the previous method, another approach was implemented. This algorithm is based on smoothing the DHRTF curve by moving average, when each sample of the transfer function is obtained as an average of its neighbors. This process can be also interpreted as convolution of the DHRTF and convolution kernel $M_A[\Omega]$, as described below.

$$\left|\bar{H}_D^{\varphi}[\Omega]\right| = \left|\widehat{H}_D^{\varphi}[\Omega]\right| * M_A[\Omega] \tag{4.13}$$

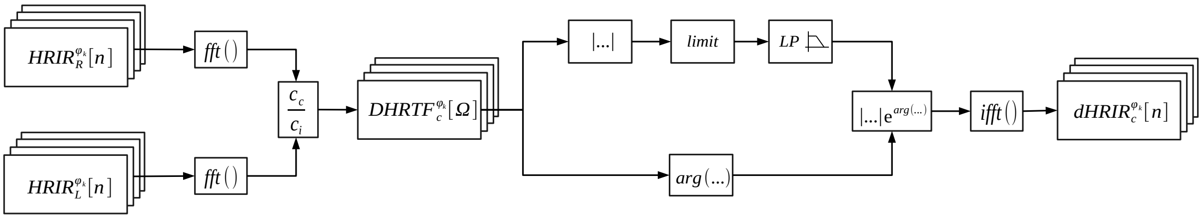

Figure 4.24: dHRIR pre-processing. A block scheme of constructing dHRIR with recommended pre-processing for the reduction of the artifact occurrence is depicted. Once the DHRTF is obtained from the HRTF pair, the phase is extracted, the module is finally limited, smoothed, and merged with the original phase. The inverse Fourier transform is the last step.

The $M_A[\Omega]$ is defined as a series of uniformly weighed coefficients of length K with amplitude of $1/K$, as shown here:

$$M_A[\Omega] = \begin{cases} 1/K & \text{for } \Omega \in (0, K) \\ 0 & \text{elsewhere.} \end{cases} \tag{4.14}$$

The $M_A[\Omega]$ can be comprehended as coefficients of impulse response of a low-pass FIR filter, therefore, Eq. 4.13 refers also to low-pass filtering as well [1]. However, this algorithm ensures that the magnitude of the DHRTF will not cross the reference level ε again.

4.3.3 Artifact Reduction in Time and Frequency Domain

This subsection focuses on the final effect of the artifact reduction both in the time and the frequency domain. For comparison of the final effect of the particular methods on the DHRTF magnitude, see Fig. 4.22. The DHRTF in the Figure was randomly chosen for demonstration purposes. The procedure of obtaining the preprocessed dHRIR ready for virtual sound source positioning is summarized in Fig. 4.24. As can be seen in the diagram in Fig. 4.24, all the processing is done within the DHRTF magnitude. The result of application of the introduced methods on a whole set of the DHRTF for positions ranging in $\varphi \in (0, 360)$ is shown in Fig. 4.23. Regarding the time domain aspects, phase extraction process is essential here. In the first step, the phase of the DHRTF is extracted as

$$\Psi_c^{\varphi}[\Omega] = arg\Big(H_D^{\varphi}[\Omega]\Big) \tag{4.15}$$

and after the processing algorithms, the original phase is returned to the signal. The phase characteristics contains the information about the ITD. For demonstration, see actual values of both HRIR's onsets and resulting ITD obtained from a set of HRIRs measured on acoustic manikin available in [17], which are shown in Fig. 3.7.

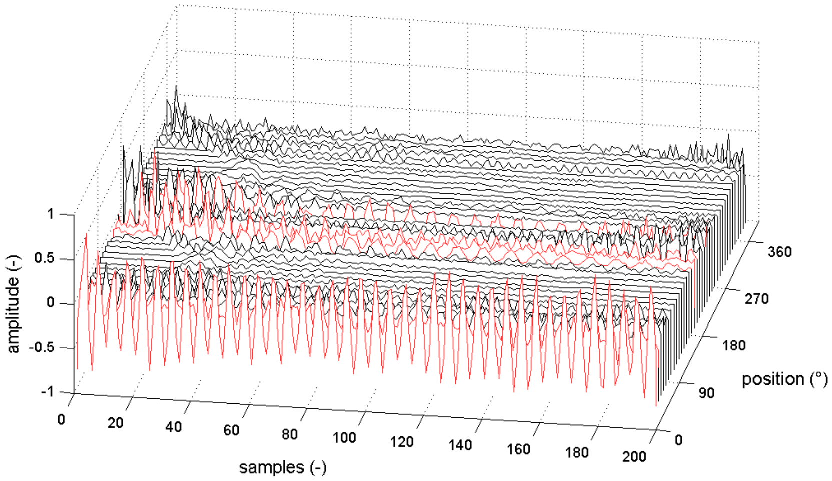

Figure 4.25: An original dHRIR set. A set of $dhrir[n]$ derived from the author's own set of HRTFs is shown (Lodz). Particular responses subjected to artifact occurrence resulting from spectral negative ILD spikes are highlighted in red color.

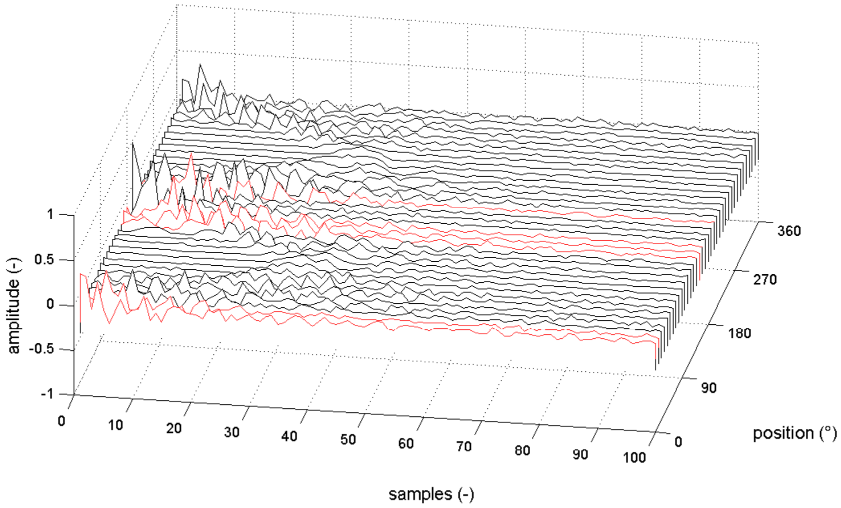

Figure 4.26: A processed dHRIR set. The effect of processing focused on artifact reduction is well observed. The problematic responses highlighted in red lose their pseudo-periodical character and extensive amplitude.

The maximum ITD usually corresponds to approximately 29-33 samples in dependence on the head proportions. The ITD is symmetrical with almost linear character. Figure 4.25 shows a pure set of dHRIRs an azimuth in range of $\varphi \in (0, 360)$ with step of 5 degrees. The dHRIR data were constructed from the author's own measured HRTF set. The particular spike-like spectral features similar to the one demonstrated in Fig. 3.12 results in pseudo-periodical character of the dHRIR. Several selected responses corresponding to heavily distorted DHRTFs are marked in red. The biggest distortion is observable for positions close to frontal axis, i.e. near $\varphi \in \{0, 180\}$.

For demonstration of the final effect of the artifact reduction algorithm on the dHRIR response, see Fig. 4.26. After application of the reduction algorithm, the pseudo-periodicity of the particular responses is suppressed as the spike-like spectral features in the DHRTF magnitude are removed. Notice also decreased noise level in the dHRIRs. Application of the low-pass filter performed by moving average also reduces the *noisy tail* of the response. Another important feature in the dHRIR set is a pair of clearly visible triangular shapes in the horizontal plane indicating the onset of the dHRIR. The shift corresponds to the ITD and the triangle profile is the same as presented in Fig. 3.7.

As stated above, the dHRIR (DHRTF) contains information about relative time shift and relative attenuation of the contra-lateral (farther) channel in relation to the ipsi-lateral (closer). The ITD is visible through the onset and the attenuation is observable through energy of the response. For positions close to the axis of $\varphi \in \{0, 180\}$, the response energy is much higher than for the side positions $\varphi \in \{90, 270\}$, where the attenuation of the contra-lateral channel is the highest.

4.3.4 Listening Tests: Artifact Reduction

The technical analysis of the particular algorithms can reveal some expectations; however, their sufficiency results from assessment of the methods by subjective listening tests. The main goal of the tests was to answer whether the introduced methods are able to remove or reduce the artifacts in the positioned sound and to rate their effectiveness. Two particular tests were performed. In the *Experiment A*, the subjects assessed the artifact reduction in dynamic-positioned sound in the whole half-plane $\varphi \in (0, 180)$. The *Experiment B* was focused on the virtual positions with higher occurrence of the artifacts ranging in $\varphi \in (0, 30)$. The details are provided below.

4.3.4.1 Experiment A: Artifacts in Half-Plane Range

Experimental setup

The listening experiment took place in a sound insulated booth. The test equipment was placed in front of the listeners' seats; computer monitor, keyboard, mouse, and reference headphones (Sennheiser HD650). The test material was played back by a computer placed outside the booth, equipped with multichannel sound card (RME Fireface UC) and headphones connected to its output. The headphones were calibrated in order to maintain the same sound pressure level (SPL) in both channels at 1 kHz. For the calibration procedure an artificial ear (Brüel & Kjær 4153) and microphone conditioner (Brüel & Kjær Nexus) connected to the computer via sound card were used. Logarithmic sweep harmonic function served as a measuring signal (see [71] for further details). Both headphone cans were measured 25 times attached on the artificial ear with added force 5 N. After each measurement the headphones were taken off and rearranged on the artificial ear in order to simulate different position of headphones on a listener's head. All the measurements were transformed into the frequency domain and the mean value of their amplitude spectra was taken for the purpose of calibration. For more information see [94].

Stimuli

Five different test materials with duration ranging from 1.7 to 2.4 s were chosen: guitar chord, snare drum phrase, white noise, singing segment, and speech segment. The selection was made to provide different spectral and temporal content within each stimuli. Since there is strong variation of the negative ILD in every HRTF set (see Sec. 1), three different HRTF set were chosen for positioning of the signal: two from the CIPIC HRTF Database (KEMAR manikin with small and large ears) [17] and one measured on the author of this thesis at Lodz University of Technology, Poland [68]. In order to involve a wider scale of employed DHRTFs, dynamic positioning was chosen for the test. The virtual position of each signal changed continuously from 0° to 180° (from front to back). Therefore, the signal was split into frames by overlap-add method using Hamming window [95]. Convolution with dHRIR corresponding to the desired virtual sound source angle was then calculated within each frame. More details regarding dynamic sound source positioning can be found in [A.4].

After the processing, loudness of each test signal was matched according to ITU-R BS.1770-2 recommendation [96] to -23 dB LUFS in Adobe Audition program environment. The parameters of the particular methods were set as follows, according to the preliminary tests: threshold level $\varepsilon = 1$, cut-off frequency (normalized) $\xi_0 = 0.4$, length of MA filter $K = 5$ at $f_s = 44.1$ kHz.

Subjects

Twenty (13 males, 7 females) listeners aged between 19 and 46 years participated in this study. Their pure tone hearing thresholds were within a range of 15 dB HL [96] for frequencies between 0.25 and 8 kHz. The subjects had no or little prior experience with this type of experiment.

Procedure

The test was performed during one session. The session consisted of a learning and a testing phase. During the learning phase three stimuli (mono reference, HRTF-positioned, and DHRTF-positioned with artifacts) were played back to the subject in order to provide him/her with sufficient information about the form of the artifact appearance in the recordings.

In the testing phase the stimuli were divided into 15 distinct groups according to the HRTF set used to obtain the DHRTF and according to the type of the testing material (guitar chord, snare drum phrase, white noise, singing, and speech segment). Every group consisted of 4 stimuli, each treated by a different technique to reduce the artifacts in the DHRTF (see Sec. 3) including the unaffected original. The testing pairs were created within each group by an adaptive square design [97] producing 60 possible pairs in total. The pair order was randomized within and also inside the pair.

After performance of each subject the pair square matrix was updated according to the scores calculated by Bradley-Terry (B-T) model [98] implemented in Matlab. New arrangement results that the next subject will compare pairs, which are similar in quality, following the adaptive square design principle [97]. For the testing phase a Matlab GUI was made. It consists of two large buttons, labeled A/B, and shown on the listener's computer screen. During the playback a color of the buttons turned green to emphasize currently played stimulus. After presenting the stimuli, the listener was asked to choose the preferred stimulus by clicking on the corresponding button. The listeners were able to replay the pair multiple times by clicking on the right mouse button.

4.3.4.2 Experiment B: Focus on Artifacts-Related Positions

Early results from the Experiment A showed little significant differences between the presented artifact reduction methods. Therefore, in Experiment B the number of virtual positions of each signal was reduced to emphasize ones with occurring artifacts. Since the artifacts occur most frequently near the front and back medial lobe (see Fig. 4.20), the DHRTFs for positions ranging between $\varphi = 0^\circ$ and $\varphi = 30^\circ$ were used. The Lodz HRTF set was chosen as it shows the highest negative ILD occurrence.

Experimental setup

The same experimental setup as in the Experiment A was used.

Stimuli

Stimuli from Experiment A were used except of the record of guitar, which was replaced by a record of saxophone. This change was motivated by the results of Experiment A.

Subjects

Eleven (7 males, 4 females) listeners aged between 21 and 46 participated in the Experiment B. Their pure tone hearing thresholds were within a range of 15 db HL for frequencies between 0.25 and 8 kHz. Three subjects also participated in Experiment A.

Procedure

The same procedure as in experiment A was employed. The only difference was that a full pair design was used instead of adaptive square design.

4.3.5 Objective Assessment

For comparison, objective assessment methods were employed in order to compare the results of the subjective tests. The PEMO-Q (see [99]) algorithm was applied to the data from Experiment B, where the artifacts were expected to occur more frequently. The original monaural stimuli were used as a reference and compared to the DHRTF-positioned samples.

4.3.6 Results

From the subjective evaluation of the Experiments A and B the pair preference matrices were obtained for each type of stimuli and the HRTF set used, thus 15 and 5 matrices, respectively. It is desirable to transform the obtained pair preference matrices into the continuous scale. For this purpose a Matlab implementation of Bradley-Terry model [98] was utilized. The output of the model (B-T score) is a maximum likelihood estimation of the scale parameters, i.e. the preference scales for the four different methods of artifact reduction with 95 % confidence intervals. The scale is logarithmic; therefore, the least preferred method has the largest absolute value. Since the B-T parameters are unique up to multiplication by a positive constant [98], they were normalized in order to achieve B-T score of -10 for unprocessed (pure) DHRTF in all cases.

The goodness of fit of the choice models was evaluated (H_0 that the BTL model holds the data). The test statistic is approximately χ^2 distributed with degrees of freedom equal to 3 in our case [98]. Bradley-Terry model also allows to determine, whether the data are statistically different from uniform [100]. The test T_U is distributed approximately χ^2 with degrees of freedom in our case equal to 3. The data is thus considered with 95 % statistical confidence non-uniform when condition $T_U > \chi^2(3)$ is fulfilled ($\chi^2(3) = 7.82$) [100].

Experiment A

The test statistics of goodness of fit and uniformity of scores of B-T model for experiment A are summarized in Table 1. It's clearly visible that the B-T model fitted all the presented data. However, test for uniformity of data T_U shows that the data for guitar (for all HRTF sets) and speech stimulus (for CIPIC KEMAR Large HRTF set) do not hold the 95 % significance condition ($T_U > 7.82$). Therefore, the data is not statistically different from a pure random sample from a uniform distribution and that the subjects were unable to perceive any audible difference between the presented methods. In Figure 4.27 the B-T scores (bars) with 95 % confidence intervals (whisker) for all stimuli and DHRTF sets (one subplot for each) used in Experiment A are shown.

Experiment B

The test statistics of goodness of fit and uniformity of scores of B-T model for experiment B are summarized in Table 2. B-T model fitted all the subjective data, also the test for uniformity T_U holds in all cases 95 % significance condition ($T_U > 7.82$). It is concluded that

	HRTF set	$\chi^2(3)$	p	T_U
Guitar	KEMAR Small	2.14	0.544	1.78*
	KEMAR Large	3.69	0.296	2.93*
	Lodz	0.35	0.840	1.14*
Noise	KEMAR Small	0.38	0.945	14.45
	KEMAR Large	0.65	0.884	18.30
	Lodz	1.92	0.589	16.48
Singing	KEMAR Small	0.61	0.895	22.82
	KEMAR Large	0.24	0.626	18.38
	Lodz	1.78	0.620	25.37
Snare	KEMAR Small	3.37	0.338	26.08
	KEMAR Large	4.22	0.239	10.92
	Lodz	0.18	0.981	24.23
Speech	KEMAR Small	1.15	0.764	20.11
	KEMAR Large	1.36	0.714	1.38*
	Lodz	1.46	0.692	8.34

Table 4.4: Experiment A: Evaluated B-T goodness of fit test statistics and corresponding p values, and results of whether the obtained data are statistically significant from uniform. Note: $*T_U < 7.82$.

	$\chi^2(3)$	p	T_U
Sax	1.17	0.760	28.9580
Noise	1.74	0.627	17.3764
Singing	2.10	0.552	25.9340
Snare	1.40	0.705	20.7524
Speech	1.63	0.654	18.3303

Table 4.5: Experiment B: Evaluated B-T goodness of fit test statistics and corresponding p values, and results of test whether the obtained data are statistically significant from uniform.

the data are statistically different from a pure random sample from a uniform distribution and there exist audible preference over the presented methods. In Figure 4.28 the B-T scores (bars) with 95 % confidence intervals (whisker) for all stimuli and DHRTF sets (one subplot for each) used in Experiment A are shown.

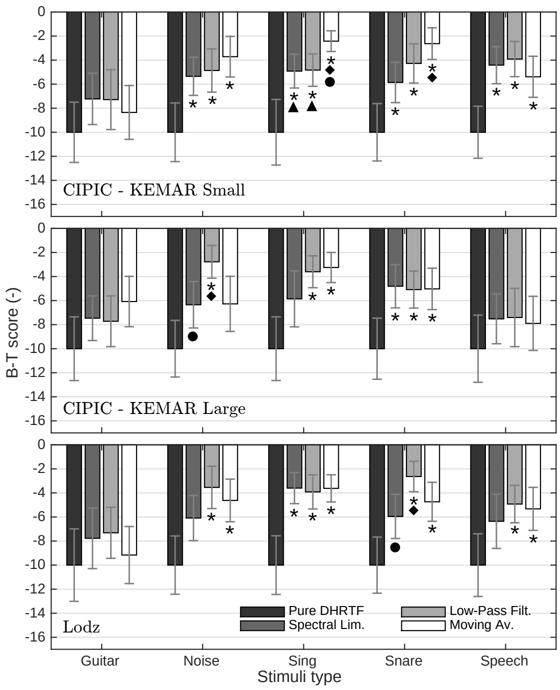

Figure 4.27: Results of the Experiment A. Each subplot corresponds to a specific HRTF set denoted in the lower left corner. The x axis represents particular stimuli, while the y axis denotes B-T score (bars). The whiskers correspond to 95 % confidence intervals. The symbol under the whisker indicates whether the method is significantly different from: ☆ – Pure DHRTF, ◇ – Spectral limitation,○ – Low-Pass filtering, △– Moving average.

Objective test

In the last step, the stimuli were compared by objective assessment method. The output of the method is represented by rate of difference to a reference sound, *Objective Difference Grade* (ODG), according to ITU-R BS.1387-1 [101]. The ODG grade is defined within an interval form 0 (no objective difference) to -4 (maximal objective difference). The results in Fig. 4.29 show maximal difference of the reference and DHRTF-positioned signal. The *hard limiting* method shows large difference as well, however, always lower than the unprocessed positioned signal. The largest similarity with the original signal occurs almost equally for moving average and low pass filtering, which corresponds to the results of the subjective tests. Note that the ODG grade should not be expected to approach 0 level since the positioned signal still contains differences caused by ITD and ILD.

4.3.7 Discussion

This section introduces three methods for artifacts reduction in the DHRTF-based positioning. The methods are intended to be used in preprocessing of the DHRTF function in

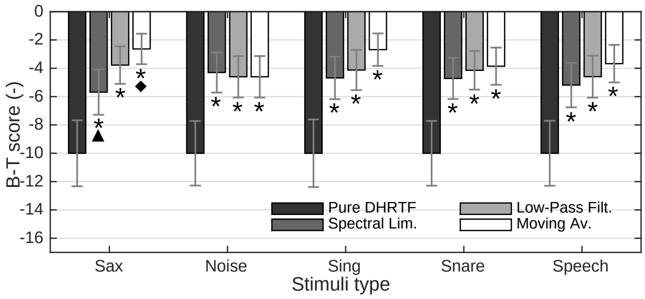

Figure 4.28: Results of the Experiment B. The results correspond to the Lodz HRTF set used to generate the DHRTF. The x axis represents particular stimuli, while the y axis denotes B-T score (bars). The whiskers correspond to 95 % confidence intervals. The ymbol under the whisker shows whether the method is significantly different from : ☆ – Pure DHRTF, ◇ – Spectral limitation, ❍ – Low-Pass filtering, △ – Moving average.

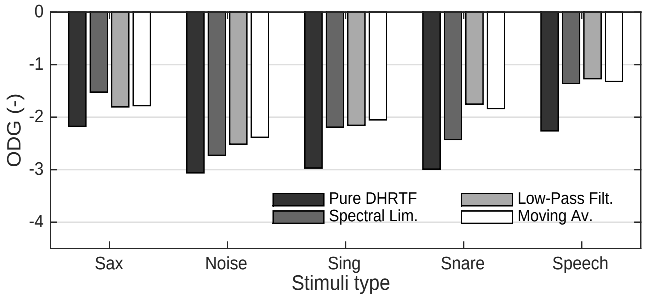

Figure 4.29: Results for Objective assessment. The results were obtained according to ITU-R BS.1116-1. The ODG grade represents a rate of difference to the reference signal, monaural non-positioned sound.

order to avoid the effect of the negative ILD occurrence while preserving all the necessary localization cues and natural timbre of the positioned sound. The subjective and objective comparisons showed that it is reasonable to employ even elementary pre-processing algorithms to improve the quality of the positioned sound.

As can be observed from the results, hard spectral limitation (Sec 4.3.2.1) provides robust basis of the artifact reduction algorithm. The B-T score increases rapidly when this method is applied (increase from 2.1 to 5.6 in particular cases) due to elimination of the spectral peaks exceeding the threshold level (1 ≈ 0 dB) in the DHRTF module. However, some other peaks can remain under the threshold level still perceptible. Its reduction by smoothing either by low-pass filtering or by moving average was chosen as the next step of preprocessing. This approach led to only small significant increase of the stimuli preference. The results of Experiment 2 focused on the *problem* positions are more consistent within the methods with the best preference of the *moving average* smoothening method, see Fig. 4.28.

In general, the results are noticeably stimuli-dependent, since the spectral content of the stimuli can affect the final form of the artifacts. For instance, the tone-like guitar stimulus shows a very small difference across the artifact reduction methods. This probably results from the discrete character of its tonal spectrum. Therefore, emphasizing the higher harmonics (timbre affection) do not need to be perceived as an unpleasant distortion in this case. Stimuli with wider range of higher frequency content (snare, singing) refers to a more distinguishable difference among the methods, since the artifacts occur in a form of well noticeable *birdies*.

The dependence on the HRTF set is primarily caused by a specific position of the negative ILD in each particular set. The cause of the spike occurrence may result from uncertainties in the HRTF measurement, especially preserving gains of the both measuring microphone equal. Nevertheless, the negative ILD occurs as well in HRTF sets rendered by the structural model [46], which is not subject to any measuring issues. The origin of the negative ILD is an object of further work. The author also plans to focus on the possibility of re-introducing the monaural spectral cues to the DHRTF by additional filtering in order to allow at least partial elevation positioning.

The results indicate that the method of spectral amplitude limiting (Sec. 4.3.2.1) significantly reduces the artifacts. Better results are then achieved by both low-pass and moving average filtering. Therefore, it is highly recommended to pre-process the DHRTF set before use. The method of *moving average* filtering was chosen by the author for pre-processing, since it can be more easily implemented compared to the *low-pass filtering* method while showing comparable results.

Chapter 5

Conclusions

This doctoral thesis focuses on one-channel virtual sound source positioning based on employment of the Differential Head-Related Transfer Function (DHRTF). The principle of the DHRTF positioning method is an extraction of frequency-dependent ITD and ILD from a pair of HRTFs and its application to the ipsi-lateral channel of a binaural signal.

5.1 Summary

Contemporary trends in the research regarding virtual sound source positioning are summarized in Chapter 2 *State-of-the-art.* Most effort is put on how to synthesize, measure, or model the HRTF and how to use virtual positioning in the field of assistive technology. Some other work is focused on front-back plane recognition or on the dynamic aspects and compensation of the head movements by a head-tracking device.

Chapter 3 *Proposed method* introduces the theoretical concept of the Differential HRTF and demonstrates the principle in time domain on the 1-D example (one response) and the 2-D example (set of responses). Spectral aspects of the DHRTF are discussed as well with highlighted circumstances for the DHRTF artifact occurrence. Theoretical design of the one-channel positioning algorithm is the main contribution of this doctoral thesis.

Chapter 4 *Main Results* is divided into three subsections, where each part represents the core and the results of particular publications dealing with the specific aspect of the DHRTF positioning. Each subsection performs listening tests focused on a specific task.

Performance of the DHRTF positioning method is investigated in Section 4.1. The DHRTF method is compared to two other common positioning methods: amplitude panning and HRTF processing. Results of technical analysis and quality assessment of the methods by subjective listening tests are presented. The tests focus on distinctive aspects of the positioning methods: spatial impression, timbre affection and loudness fluctuations. Subjective tests have shown that the proposed DHRTF positioning method shows promising and statistically significant results in comparison with the other widely used methods of virtual sound source positioning: amplitude panning and HRTF positioning. Due to its one-channel filtering, the DHRTF can be applied in devices and setups with limited access to computational resources. The results show that the DHRTF positioning method is applicable with very promising performance; it avoids perceptible channel coloration that occurs within the HRTF method, and it delivers spatial impression more successfully than the simple amplitude panning method. The listening tests discovered that an important advantage of the DHRTF method is the preservation of the original sound timbre, which may be utilized in musical applications requiring separation of the sources in the stereo base (e.g. common mixing procedure in song production) for aesthetic purposes. Such mixing procedure would deliver more natural spatial separation of the sources (instruments) than the commonly used amplitude panning, while not affecting timbre of particular tracks as when the HRTF method is employed.

Section 4.2 focuses on the aspects of implementation of moving a virtual sound source in the virtual auditory environment. Rendering of the dynamic source by use of (Differential) Head-Related Transfer Function is achieved by filtering with varying impulse responses. The main difference results from the different output of linear and circular convolutions usually implemented in such processing. Particular variants are described, discussed and compared by subjective listening tests of the positioned signals. For proper implementation of the moving virtual source created by the set of (D)HRTFs, linear convolution is necessary to be implemented. If this requirement is not fulfilled, the occurring inter-channel differences, normally indistinguishable under mono-listening conditions, will weaken the final spatial impression for most of the cases. This is primarily caused by absence of the proper ITD, a crucial factor for localization cues. Due to processing artifact occurrences, it is also recommended to involve overlap-add techniques in the implementation algorithm. This research used overlapping factor o = 0.5 sufficiently. Length of L segment is recommended in the range of 1024 to 4096 samples, ideally 2048. This step will help to avoid processing artifacts and allow sufficient angular frequency of the source without significant leaps in source position.

In Section 4.3, specific methods for reducing artifacts, which may occur within the DHRTF-based virtual sound source positioning method employing the DHRTF, are introduced. The artifacts result from a process of obtaining the DHRTF, where spike-like spectral peak(s) may occur in the module of the DHRTF. These spectral features result in an undesirable whistling-like sound component that distorts both timbre and spatial perception of the virtual sound. The reduction methods are intended to be used in preprocessing of the DHRTF function in order to avoid the effect of the negative ILD occurrence while preserving all the necessary localization cues and natural timbre of the positioned sound. Their performance is assessed by both subjective and objective tests. Three selected methods based on the limitation and smoothening of the DHRTF magnitude by a low-pass filter are introduced and compared by both paired comparison listening tests and the objective assessment method. The tests showed significant reduction of the artifact occurrence. The best results were obtained for the method based on limiting the DHRTF magnitude and its smoothening by a moving average convolution kernel.

5.2 Contributions of the Thesis

This thesis provides several original contributions and results presented throughout the text. The results relevant to the thesis scope were also published in multiple articles including two journals with the impact factor ([A.1], [A.2]). The highlights of the thesis are specifically as follows:

1. A new positioning algorithm enabling reduction of virtual positioning process from usual double-channel processing to only one-channel processing was designed. The core of this algorithm is the *Differential Head-Related Transfer Function* (DHRTF).

2. Theoretical principles of the DHRTF method were introduced and acquisition of the DHRTF from an existing pair of HRTFs were presented.

3. The thesis describes the detailed implementation of the DHRTF algorithm and compares it to common positioning algorithms for static and dynamic virtual positioning.

4. Multiple specific listening tests are used to compare the DHRTF algorithm to the common positioning methods in terms of quality of spatial perception and timbre character of the final processed sound were designed, performed, and evaluated.

5. Artifacts, which may occur in the sound virtually positioned by the DHRTF method as a result of specific processing features, were explored. Methods for their elimination were designed and verified by listening tests.

6. The basics of virtual sound source positioning and its state-of-the-art was summarized into an integrated survey introduced at the beginning of the doctoral thesis.

5.3 Future Work

Since the thesis presents primarily published results (see Publications of the Author), preliminary results and potential directions of further exploration were excluded from the content. A very brief summary of what the author suggests to explore is as follows:

- Concluded from the results in Section 4.3, it is very challenging to investigate further origins of the Negative ILD. Current results indicate that the NILD occurs primarily in the measured HRTF sets when gains of both measuring microphones are not equal. However, preliminary results obtained during the thesis editing discovered that the NILD is present even in particular analytical models for the HRTF synthesis.

- Other listening tests are worth considering in order to assess the DHRTF performance with multiple positioned sources. Comparison to the HRTF and the AP would be beneficial here.

- The implementation of the DHRTF can be improved by employing a Head-Tracking device and Doppler frequency shift. It would be interesting to explore, how the method performs during more realistic listening conditions.
- The DHRTF would be beneficial as a VST plugin for digital interfaces for mixing music. Less channel coloration (versus the HRTF method) may provide impressive results.
- As discussed in Sec. 4.1.1.2, certain comprehension of the DHRTF expresses inter-channel coloration. Further exploration would be beneficial to see how perception of different coloration switches to spatialization of the sound.
- The author also plans to focus on re-introducing monaural spectral cues to the DHRTF by implementing additional filtering in order to achieve partial elevation positioning.

References

[1] Udo Zölzer. *DAFX: Digital Audio Effects*. Wiley Online Library, Hoboken, NJ, USA, 2011.

[2] Jens Blauert. *The technology of binaural listening*. Springer Verlag, Berlin, 2013. eBook.

[3] Barbara G Shinn-Cunningham, Scott Santarelli, and Norbert Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. *Journal of the Acoustic Society of America*, 107(3):1627–1636, 2000.

[4] Yiteng Huang and Jacob Benesty. *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Springer, Boston, MA, USA, 2004.

[5] Shu-Nung Yao and Li Jen Chen. HRTF adjustments with audio quality assessments. *Archives of Acoustics*, 38(1):55–62, 2013.

[6] Norman H Adams and Gregory H Wakefield. State-space synthesis of virtual auditory space. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(5):881–890, 2008.

[7] V Ralph Algazi and Richard O Duda. Headphone-based spatial sound. *Signal Processing Magazine, IEEE*, 28(1):33–42, 2011.

[8] Francis Rumsey. Whose head is it anyway? Optimizing binaural audio. *The Journal of the Audio Engineering Society*, 59(9):672–675, 2011.

[9] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[10] Richard O Duda. Auditory localization demonstrations. *Acta Acustica United with Acustica*, 82(2):346–355, 1996.

[11] C Phillip Brown and Richard O Duda. An efficient HRTF model for 3-D sound. In *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*, pages 4–pp. IEEE, 1997.

[12] Richard O Duda. Modeling head related transfer functions. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 996–1000. IEEE, 1993.

[13] Allen William Mills. On the minimum audible angle. *Journal of the Acoustic Society of America*, 30:237–246, 1958.

[14] Janina Fels and Michael Vorländer. Anthropometric parameters influencing head-related transfer functions. *Acta Acustica united with Acustica*, 95(2):331–342, 2009.

[15] V Ralph Algazi and Richard O Duda. Effective use of psychoacoustics in motion-tracked binaural audio. In *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, pages 562–567. IEEE, 2008.

[16] William Morris Hartmann and Brad Rakerd. On the minimum audible angle - a decision theory approach. *Journal of the Acoustic Society of America*, 85(5):2031–2041, 1989.

[17] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. The CIPIC HRTF database. In *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.

[18] Ewan A Macpherson and John C Middlebrooks. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *Journal of the Acoustic Society of America*, 111(5):2219–2236, 2002.

[19] Petr Marsalek. Neural code for sound localization at low frequencies. *Neurocomputing*, 38:1443–1452, 2001.

[20] Petr Marsalek and Jiri Kofranek. Sound localization at high frequencies and across the frequency range. *Neurocomputing*, 58:999–1006, 2004.

[21] Lubomir Kostal and Petr Marsalek. Neuronal jitter: can we measure the spike timing dispersion differently. *Chinese Journal of Physiology.*, 53:454–464, 2010.

[22] Pavel Sanda and Petr Marsalek. Stochastic interpolation model of the medial superior olive neural circuit. *Brain Research*, 1434:257–265, 2012.

[23] Tianyi Yan and Jinglong Wu. The contribution of pinna to the discriminate the vertical angle for virtual reality technology. In *SICE, 2007 Annual Conference*, pages 3080–3083. IEEE, 2007.

[24] Robert AA Campbell, Andrew J King, Fernando R Nodal, Jan WH Schnupp, Simon Carlile, and Timothy P Doubell. Virtual adult ears reveal the roles of acoustical factors and experience in auditory space map development. *The Journal of Neuroscience*, 28(45):11557–11570, 2008.

[25] Erno HA Langendijk and Adelbert W Bronkhorst. Contribution of spectral cues to human sound localization. *Journal of the Acoustic Society of America*, 112(4):1583–1596, 2002.

[26] Elena Blanco-Martin, Francisco Javier Casajús-Quirós, Juan José Gómez-Alfageme, and Luis Ignacio Ortiz-Berenguer. Objective measurement of sound event localization in horizontal and median planes. *The Journal of the Audio Engineering Society*, 59(3):124–136, 2011.

[27] ES Malinina and IG Andreeva. The role of spectral components of the head-related transfer functions in evaluation of the virtual sound source motion in the vertical plane. *Acoustical Physics*, 56(4):576–583, 2010.

[28] Carlos Avendano, V Ralph Algazi, and Richard O Duda. A head-and-torso model for low-frequency binaural elevation effects. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pages 179–182. IEEE, 1999.

[29] V Ralph Algazi, Richard O Duda, Reed P Morrison, and Dennis M Thompson. Structural composition and decomposition of HRTFs. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 103–106. IEEE, 2001.

[30] V Ralph Algazi, Carlos Avendano, and Richard O Duda. Elevation localization and head-related transfer function analysis at low frequencies. *Journal of the Acoustic Society of America*, 109:1110, 2001.

[31] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. Fitting pinna-related transfer functions to anthropometry for binaural sound rendering. In *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pages 194–199. IEEE, 2010.

[32] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Estimation and modeling of pinna-related transfer functions. In *Proceedings of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, pages 6–10, 2010.

[33] Navarun Gupta, Armando Barreto, and Carlos Ordonez. Improving sound spatialization by modifying head-related transfer functions to emulate protruding pinnae. In *SoutheastCon, 2002. Proceedings IEEE*, pages 446–450. IEEE, 2002.

[34] Qiong Yuan and Nai-Hwa Chiang. *An exploratory study of people with vision impairments adjusting to new environments.* DePaul University, School of CDM, 2010.

[35] Monika Rychtáriková, Tim Van den Bogaert, Gerrit Vermeir, and Jan Wouters. Binaural sound source localization in real and virtual rooms. *The Journal of the Audio Engineering Society*, 57(4):205–220, 2009.

[36] Matti Gröhn. Localization of a moving virtual sound source in a virtual room, the effect of a distracting auditory stimulus. In *International Conference on Auditory Display, Kyoto, Japan.* Citeseer, 2002.

[37] Tapio Lokki, Matti Grohn, Lauri Savioja, and Tapio Takala. A case study of auditory navigation in virtual acoustic environments. In *Proceedings of International Conference on Auditory Display (ICAD2000).* Citeseer, 2000.

[38] U Peter Svensson. Modelling acoustic spaces for audio virtual reality. In *Proceedings of the IEEE Benelux Workshop on Model Based Processing and Coding of Audio*, pages 109–116, 2002.

[39] Simon Carlile, Craig Jin, and Vaughn Harvey. The generation and validation of high fidelity virtual auditory space. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 3, pages 1090–1095. IEEE, 1998.

[40] William G Gardner. *3-D audio using loudspeakers.* Springer Science & Business Media, 1998.

[41] V Ralph Algazi, Eric J Angel, and Richard O Duda. On the design of canonical sound localization environments. In *Audio Engineering Society Convention 113.* Audio Engineering Society, 2002.

[42] Tapio Lokki, Lauri Savioja, Riitta Vaananen, et al. Creating interactive virtual auditory environments. *IEEE Computer Graphics and Applications*, 22:49–57, 2002.

[43] Ramani Duraiswami, Dmitry N Zotkin, Zhiyun Li, Elena Grassi, Nail A Gumerov, and Larry S Davis. High order spatial audio capture and binaural head-tracked playback over headphones with HRTF cues. In *Audio Engineering Society Convention 119*, 2005.

[44] Ville Pulkki. Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning. *The Journal of the Audio Engineering Society*, 49(9):753–767, 2001.

[45] William M Hartmann and Andrew Wittenberg. On the externalization of sound images. *Journal of the Acoustic Society of America*, 99(6):3678–3688, 1996.

[46] C Phillip Brown and Richard O Duda. A structural model for binaural sound synthesis. *Speech and Audio Processing, IEEE Transactions on*, 6(5):476–488, 1998.

[47] Richard O Duda, Carlos Avendano, and V Ralph Algazi. An adaptable ellipsoidal head model for the interaural time difference. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 2, pages 965–968. IEEE, 1999.

[48] Alan V Oppenheim, Ronald W Schafer, John R Buck, et al. *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs, 1989.

[49] Corey I Cheng and Gregory H Wakefield. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. In *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[50] Jaka Sodnik, Rudolf Susnik, and Saso Tomazic. Acoustic signal localization through the use of head related transfer functions. *Systemics, Cybernetics and Informatics*, 2(6):56–59, 2004.

[51] Patrick Roth, Lori Petrucci, André Assimacopoulos, and Thierry Pun. AB-Web: Active audio browser for visually impaired and blind users. In *Proceedings of International Conference on Auditory Display*, volume 98, 1998.

[52] Patrick Roth, Lori Stefano Petrucci, André Assimacopoulos, and Thierry Pun. Audio-haptic internet browser and associated tools for blind and visually impaired computer users. In *Workshop on friendly exchanging through the net*, pages 22–24. Citeseer, 2000.

[53] Mansoor Hyder, Michael Haun, and Christian Hoene. Placing the participants of a spatial audio conference call. In *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, pages 1–7. IEEE, 2010.

[54] Dmitry N Zotkin, Ramani Duraiswami, and Larry S Davis. Rendering localized spatial audio in a virtual auditory space. *Multimedia, IEEE Transactions on*, 6(4):553–564, 2004.

[55] Udo Zölzer, Xavier Amatriain, and John Wiley. *DAFX: digital audio effects*, volume 1. Wiley Online Library, 2002.

[56] Andreas Antoniou. *Digital signal processing*. McGraw-Hill Toronto, Canada:, 2006.

[57] V Ralph Algazi, Richard O Duda, and Dennis M Thompson. Motion-tracked binaural sound. *The Journal of the Audio Engineering Society*, 52(11):1142–1156, 2004.

[58] Michal Bujacz, Piotr Skulimowski, and Pawel Strumillo. Navitona prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired. *The Journal of the Audio Engineering Society*, 60(9):696–708, 2012.

[59] Yinlin Li, Christoph Groenegress, Jochen Denzinger, Wolfgang Strauss, and Monika Fleischmann. An acoustic interface for triggering actions in virtual environments. In *Fourth international conference On Virtual Reality and Its Applications in Industry*, pages 246–251. International Society for Optics and Photonics, 2004.

[60] Yoshikazu Seki and Tetsuji Sato. A training system of orientation and mobility for blind people using acoustic virtual reality. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 19(1):95–104, 2011.

[61] Areti Andreopoulou, Agnieszka Rogińska, and Hariharan Mohanraj. A database of repeated head-related transfer function measurements. 2013.

[62] V Ralph Algazi, Pierre L Divenyi, VA Martinez, and Richard O Duda. Subject dependent transfer functions in spatial hearing. In *Circuits and Systems, 1997. Proceedings of the 40th Midwest Symposium on*, volume 2, pages 877–880. IEEE, 1997.

[63] Dmitry N Zotkin, Jane Hwang, R Duraiswaini, and Larry S Davis. HRTF personalization using anthropometric measurements. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 157–160. Ieee, 2003.

[64] Michał Pec, Michał Bujacz, and Paweł Strumiłło. Personalized head related transfer function measurement and verification through sound localization resolution. In *Proceedings of the 15th European Signal Processing Conference*, pages 2326–2330, 2007.

[65] Hongmei Hu, Lin Zhou, Hao Ma, and Zhenyang Wu. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, 69(2):163–172, 2008.

[66] Elizabeth M Wenzel, Marianne Arruda, Doris J Kistler, and Frederic L Wightman. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustic Society of America*, 94:111, 1993.

[67] Michał Pec, Michał Bujacz, P Strumillo, and Andrzej Materka. Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind. In *Signals and Electronic Systems, 2008. ICSES'08. International Conference on*, pages 235–238. IEEE, 2008.

[68] Andrzej Dobrucki, Przemyslaw Plaskota, Piotr Pruchnicki, Michal Pec, Michal Bujacz, and Pawel Strumillo. Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals. *The Journal of the Audio Engineering Society*, 58(9):724–738, 2010.

[69] Nail A Gumerov, Adam E ODonovan, Ramani Duraiswami, and Dmitry N Zotkin. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *Journal of the Acoustic Society of America*, 127:370, 2010.

[70] Dumidu S Talagala and Thushara D Abhayapala. Novel head related transfer function model for sound source localisation. In *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on*, pages 1–6. IEEE, 2010.

[71] Angelo Farina. Advancements in Impulse Response Measurements by Sine Sweeps, 2007.

[72] Zdenek Otcenasek. *On Subjective Evaluation of Sound (In Czech)*. Akademie muzickych umeni, Prague, Czech Republic, 2008.

[73] Norman H Adams. *A model of head-related transfer functions based on a state-space analysis*. Disertation thesis, The University of Michigan, 2008.

[74] Robert Baumgartner, Piotr Majdak, and Bernhard Laback. Modeling sound-source localization in sagittal planes for human listeners. *Journal of the Acoustic Society of America*, 136(2):791–802, 2014.

[75] V Ralph Algazi, Carlos Avendano, and Richard O Duda. Estimation of a spherical-head model from anthropometry. *The Journal of the Audio Engineering Society*, 49(6):472–479, 2001.

[76] D Wesley Grantham, Joel Andrew Willhite, Kenneth D Frampton, and Daniel H Ashmead. Reduced order modeling of head related impulse responses for virtual acoustic displays. *Journal of the Acoustic Society of America*, 117:3116, 2005.

[77] Peter Fiala, Jacobus Huijssen, Bert Pluymers, Raphael Hallez, and Wim Desmet. Fast multipole BEM modeling of head related transfer functions of a dummy head and torso. In *ISMA 2010 Conference, Leuven, Belgium*, 2010.

[78] John C Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *Journal of the Acoustic Society of America*, 106:1493, 1999.

[79] Carlos Avendano, Richard O Duda, and V Ralph Algazi. Modeling the contralateral HRTF. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*. Audio Engineering Society, 1999.

[80] José Luis González-Mora, A Rodriguez-Hernandez, E Burunat, F Martin, and MA Castellano. Seeing the world by hearing: Virtual acoustic space (VAS) a new space perception system for blind people. In *Information and Communication Technologies, 2006. ICTTA'06. 2nd*, volume 1, pages 837–842. IEEE, 2006.

[81] Zoltan Haraszy, David-George Cristea, Virgil Tiponut, and Titus Slavici. Improved head related transfer function generation and testing for acoustic virtual reality development. In *7th WSEAS International Conference on Engineering Education*, pages 22–24, 2010.

[82] György Wersényi. Localization in a HRTF-based minimum-audible-angle listening test for GUIB applications. *Electronic Journal of Technical Acoustics*, 1:16, 2007.

[83] György Wersényi. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(2):247–252, 2009.

[84] Peter Xinya Zhang and William M Hartmann. On the ability of human listeners to distinguish between front and back. *Hearing research*, 260(1):30–46, 2010.

[85] Vladimir Ortega-González, Samir Garbaya, and Frédéric Merienne. Reducing reversal errors in localizing the source of sound in virtual environment without head tracking. In *Haptic and Audio Interaction Design*, pages 85–96. Springer, 2010.

[86] Gaëtan Lorho, Jyri Huopaniemi, Nick Zacharov, and David Isherwood. Efficient HRTF synthesis using an interaural transfer function model. In *Signal Processing Conference, 2000 10th European*, pages 1–4. IEEE, 2000.

[87] Piotr Majdak, Matthew J Goupell, and Bernhard Laback. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics*, 72(2):454–469, 2010.

[88] Micah T Taylor, Anish Chandak, Lakulish Antani, and Dinesh Manocha. Resound: interactive sound rendering for dynamic virtual environments. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 271–280. ACM, 2009.

[89] Sen M Kuo, Bob H Lee, and Wenshun Tian. *Real-Time Digital Signal Processing: Fundamentals, Implementations and Applications.* John Wiley & Sons, 2013.

[90] Kosuke Tsujino, Wataru Kobayashi, Takao Onoye, and Yukihiro Nakamura. Efficient 3-D sound movement with time-varying IIR filters. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 90(3):618–625, 2007.

[91] Makoto Otani and Tatsuya Hirahara. A dynamic virtual auditory display: Its design, performance, and problems in HRTF switching. In *Proceedings of the Japan–China Joint Conference of Acoustics*, 2007.

[92] Jesper Sandvad. Dynamic aspects of auditory virtual environments. In *Audio Engineering Society Convention 100.* Audio Engineering Society, 1996.

[93] Yukio Iwaya and Yôiti Suzuki. Rendering moving sound with the doppler effect in sound space. *Applied Acoustics*, 68(8):916–922, 2007.

[94] Jaroslav Bouse. Headphone measurement tool implemented in Matlab. In *Proceedings of 19th International Scientific Student Conference POSTER 2015*, pages 1–5. Czech Technical University in Prague, 2015.

[95] Marina Bosi and Richard E. Goldberg. *Introduction to Digital Audio Coding and Standards.* Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[96] ITU-R. Recommendation ITU-R BS.1770-2 (Algorithms to measure audio programme loudness and true-peak audio), 2011.

[97] Jing Li, Marcus Barkowsky, and Patrick Le Callet. Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs. *Proceedings of SPIE, Stereoscopic Displays and Applications XXIV*, 8648, 2013.

[98] Florian Wickelmaier and Christian Schmid. A Matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, & Computers*, 36(1):29–40, 2004.

[99] Rainer Huber and Birger Kollmeier. PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(6):1902–1911, 2006.

[100] John C. Handley. Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment. In *Proceedings IS and Ts Image Processing, Image Quality, Image Capture, Systems Conference*, pages 108–112, 2001.

[101] ITU-R. Recommendation ITU-R BS.1387-1 (Method for objective measurements of perceived audio quality), 2001.

Publications of the Author

Journals with Impact Factor:

[A.1] Štorek, D. - Bouše, J. - Rund, F. - Maršálek, P. *Artifact Reduction in Positioning Algorithm Using Differential HRTF.* In: Journal of Audio Engineering Society, Issue 64, p. 208-217, 2016.
ISSN 1549-4950. Shares: **35**/35/20/10

[A.2] Štorek, D. - Rund, F. - Maršálek, P. *Subjective Evaluation of Three Headphone-Based Virtual Sound Source Positioning Methods Including Differential Head-Related Transfer Function.* In: Archives of Acoustics, vol. 41, no. III, p. 437-447, 2016.
ISSN 0137-5075. Shares: **34**/33/33

Reviewed Journals:

[A.3] Štorek, D. - Rund, F. - Vítek, S. - Baráth, T. *Virtual Sound Source Positioning for Navigation of Visually Impaired.* In: Ingenium. Revista de la Facultad de Ingeniera. vol. 14, no. 27, p. 6-14, 2013.
ISSN 0124-7492. Shares: **45**/35/10/10

Indexed in ISI:

[A.4] Štorek, D. *Rendering Moving Sound Source for Headphone-Based Virtual Acoustic Reality: Aspects of Signal Processing Implementation.* In: Proceedings of the 19th International Conference on Applied Electronics 2014. Pilsen, University of West Bohemia, p. 271-276, 2014.
ISBN 978-80-261-0276-2. Shares: **100**

[A.5] Štorek, D. *Virtual Sound Source Positioning by Differential Head Related Transfer Function.* In: Proceedings of the AES 49th International Conference: Audio for Games. London: Audio Engineering Society, p. 1-6, 2013.
ISBN 978-0-937803-90-5. Shares: **100**

[A.6] Štorek, D. - Rund, F. *Differential Head Related Transfer Function as a New Approach to Virtual Sound Source Positioning.* In: Proceedings of 22nd International Conference Radioelektronika 2012. Brno: VUT v Brně, FEKT, Ústav radioelektroniky, vol. 1, p. 71-74, 2012.
ISBN 978-80-214-4468-3. Shares: **70**/30

[A.7] Kadlec, F. - Rund, F. - Štorek, D. *Measurement and Analysis of Electro-Acoustic Systems for Assistive Technology.* In: 18th International Congress on Sound and Vibration. Auburn: International Institute of Acoustics and Vibration, p. 1-8, 2011.
ISBN 978-85-63243-01-0. Shares: 50/25/**25**

[A.8] Štorek, D. - Rund, F. - Suchan, R. *Virtual Auditory Space for the Visually Impaired - Experimental Background.* In: 2011 International Conference on Applied Electronics. Plzeň: Západočeská univerzita v Plzni, p. 371-374, 2011.
ISSN 1803-7232, ISBN 978-80-7043-987-6. Shares: **50**/45/5

Other Relevant Publications:

[A.9] Štorek, D. *Time Domain Aspects of Artifact Reduction in Positioning Algorithm using Differential Head-Related Transfer Function.* In: Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16), Brno, p. 191-194, 2016.
ISSN: 2413-6700. Shares: **100**

[A.10] Štorek, D. - Stuchlík, J. - Rund, F. *Modifications of the Surrounding Auditory Space by Augmented Reality Audio: Introduction to Warped Acoustic Reality.* In: ICAD 2015, Proceedings of the 21st International Conference on Auditory Display Graz: University of Graz, p. 225-230, 2015.
ISBN 978-3-902949-01-1. Shares: **40**/30/30

[A.11] Štorek, D. - Rund, F. - Baráth, T. - Vítek, S. *Virtual Auditory Space for Visually Impaired - Methods for Testing Virtual Sound Source Localization.* In: ICAD 2013 Proceedings of the International Conference on Auditory Display Lodz: Technical University of Lodz, p. 33-36, 2013.
ISBN 978-83-7283-546-8. Shares: **40**/35/15/10

[A.12] Rund, F. - Štorek, D. *Head-Tracking for Virtually Auditory Space Implemented in MATLAB.* In: 20th Annual Conference Proceeding's Technical Computing Bratislava 2012. Prague: HUMUSOFT, p. 1-4, 2012.
ISBN 978-80-970519-4-5. Shares: 70/**30**

[A.13] Štorek, D. - Rund, F. *The Graphic Analysis and Pre-processing of Differential Head Related Transfer Function.* In: 20th Annual Conference Proceeding's Technical Computing Bratislava 2012. Prague: HUMUSOFT, p. 1-6, 2012.
ISBN 978-80-970519-4-5. Shares: **90**/10

[A.14] Štorek, D. - Suchan, R. *Aspects of Influence of the Microphone Attachment in HRTF Measuring.* In: POSTER 2012 - 16th International Student Conference on Electrical Engineering. Prague: Czech Technical University in Prague, p. 1-4, 2012.
ISBN 978-80-01-05043-9. Shares: **99**/1

[A.15] Rund, F. - Štorek, D. - Glaser, O. - Barda, M. *Comparing orientation in simple HRTF-based virtual auditory space between sighted and visually impaired.* In: Posterus.sk [online: http://www.posterus.sk/?p=9798], vol. 4, no. 1, 2011.
ISSN 1338-0087. Shares: 49/**49**/1/1

[A.16] Rund, F. - Štorek, D. - Glaser, O. - Barda, M. *Using of the Sound Sources Virtual Positioning Methods for Three-dimensional Orientation of the Visually Impaired.* In: Workshop 2011. Prague: Czech Technical University in Prague, p. 1-5,
2011. Shares: 40/**30**/23/7

[A.17] Štorek, D. *Graphical User Interface for Measuring the Just Noticeable Difference in Localization of Virtual Acoustic Sources.* In: 19th Annual Conference Proceedings Technical Computing Prague 2011. Vydavatelství VŠCHT Praha, p. 1-4, 2011.
ISBN 978-80-7080-794-1. Shares: **100**

[A.18] Štorek, D. - Suchan, R. *Analysis of the Problems in HRTF Measuring.* In: POSTER 2011 - 15th International Student Conference on Electrical Engineering. Prague: CTU, Faculty of Electrical Engineering, p. 1-4, 2011.
ISBN 978-80-01-04806-1. Shares: **80**/20

[A.19] Hadrava, J. - Bernhauerm D. - Štorek, D. *The Audio Component of a Virtual Reality.* In: Proceedings of 17th Conference of Czech and Slovak Physicists. Zilina: Slovakia, 2011.
ISBN: 978-809706254-5. Shares: 34/33/**33**

[A.20] Rund, F. - Štorek, D. - Glaser, O. *GUI for Comparing Perception of Sound Adjusted by Measured or Modeled HRTF.* In: Technical Computing Bratislava 2010. Bratislava: RT systems, s.r.o, p. 1-5, 2010.
ISBN 978-80-970519-0-7. Shares: 40/**40**/20

[A.21] Rund, F. - Štorek, D. - Glaser, O. - Barda, M. *Orientation in Simple Virtual Auditory Space Created with Measured HRTF.* In: Technical Computing Bratislava 2010. Bratislava: RT systems, s.r.o, p. 1-7, 2010.
ISBN 978-80-970519-0-7. Shares: 40/**40**/10/10

[A.22] Štorek, D. - Glaser, O. *Virtual Auditory Space for Visually Impaired Created with HRTF.* In: POSTER 2010 - Proceedings of the 14th International Conference on Electrical Engineering. Prague: Czech Technical University in Prague, Faculty of Electrical Engineering, p. 1-4, 2010.
ISBN 978-80-01-04544-2. Shares: **50**/50

[A.23] Rund, F. - Glaser, O. - Štorek, D. *GUI pro demonstraci principu binaurální lokalizace zdroj zvuku.* In: Proceedings of Technical Computing Prague 2009. Prague: Humusoft, p. 1-5, 2009.
ISBN 978-80-7080-733-0. Shares: 40/30/**30**

Appendix A

Selected hard copies of this thesis contain a CD with an electronic version (pdf file) of this work with both articles [A.1] and [A.2] published in the impact factor journals.