



HAL
open science

Reconstruction de l'environnement sonore sous une protection auditive : principe et concept pour la protection auditive transparente acoustiquement

Lorenz Kroener

► To cite this version:

Lorenz Kroener. Reconstruction de l'environnement sonore sous une protection auditive : principe et concept pour la protection auditive transparente acoustiquement. Acoustique [physics.class-ph]. HESAM Université, 2021. Français. NNT : 2021HESAC009 . tel-03505926

HAL Id: tel-03505926

<https://theses.hal.science/tel-03505926v1>

Submitted on 1 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE SMI

Laboratoire de Mécanique des Structures et des Systèmes Couplés

THÈSE

présentée par : **Lorenz KROENER**

soutenue le : **6 Octobre 2021**

pour obtenir le grade de : **Docteur d'HESAM Université**

préparée au : **Conservatoire national des arts et métiers**

Discipline : **Sciences pour l'Ingénieur**

Spécialité : **Acoustique**

Reconstruction of the Audio Environment with a Hearing Protection Device: Principle and Concepts of Solutions for an Acoustically “Transparent” Hearing Protection

THÈSE dirigée par :

Monsieur Alexandre GARCIA Professeur des Universités, LMSSC, CNAM Paris

Jury:

Mme. Véronique ZIMPFER	Ingénieur de recherche HDR, APC, ISL, Saint-Louis	Co-encadrante
M. Christophe LANGRENNE	Ingénieur de recherche, LMSSC, CNAM, Paris	Co-encadrant
M. Joël DUCOURNEAU	Professeur, LEMETA , Univ. Lorraine, Nancy	Rapporteur
M. François ALOUGES	Professeur, CMAP, École Polytechnique, Paris	Rapporteur, Président du jury
Mme. Rozenn NICOL	Ingénieur de Recherche HDR, Orange Lab, Lannion	Examinatrice
M. Guillaume ANDEOL	HDR, IRBA, Paris	Examineur
M. Udo ZÖLZER	Univ.-Prof. Dr.-Ing. habil., ET/ANT, HSU, Hamburg	Examineur

**T
H
È
S
E**

Acknowledgments

The presented work of this thesis was carried out as part of a collaboration between the Battlefields Acoustics and Protection of the Soldier (APC) team of the French-German Research Institute of Saint-Louis (ISL) and the Structural Mechanics and Coupled Systems Laboratory (LMSSC) of the Conservatoire national des arts et métiers (CNAM).

First of all, I would like to express my gratitude to my on-site supervisor Dr. Véronique Zimpfer and my Paris-based supervisor Prof. Alexandre Garica. They have guided me throughout this three and a half years of PhD studies with excellent advice and lot of patience. Thanks to our fruitful discussions and their suggestions, questions, and comments, they encouraged me all along my PhD studies. Further, Christophe Langrenne co-supervised this thesis from Paris and provide me with his numerical simulation tools.

I would like to thank my team manager Pierre Naz, who always had an open ear for my administrative requests and to my colleges from the APC team, who supported me with their skills and knowledge, especially Thomas and Cyril who patiently reviewed my writings, Alexandre and Laurent from the workshop who did many manual tasks, but also Marcus and Franck, both experts in electronics, who were always available for my questions and answered in great detail. With their help, I was able to mount the loudspeaker array and manufacture the prototypes.

I would like to acknowledge and thank all the participants of my listening tests. Their willingness to participating in the experiments and not getting tired being invited several times allowed me to acquire a profound data set on which I was able to evaluate my work.

Finally, I like to thank my family who made it possible for me to do my studies and supported me all along my educational course.

ACKNOWLEDGMENTS

Abstract

Many professionals, e.g. mining and construction workers, ground crews or soldiers are exposed to impulsive or constant high level noise. In order to prevent hearing loss, they depend on hearing protection devices (HPDs). On the contrary, HPDs interfere with situational awareness and sound source localization. This contradiction makes users pondering between hearing loss and situational awareness. Often last mentioned dominates over first mentioned. This work aims to bring in line hearing protection and situational awareness. A virtual acoustic environment (VAE) with 16 circularly, horizontally arranged loudspeakers is set up. Localization performance with commercially available HPDs, including active and passive earplugs and earmuffs, is assessed in the VAE with 40 subjects. Earplugs with small geometries show better results than large-sized earmuffs. These results coincide with the study on modifications of the Head Related Transfer Function (HRTF) caused by HPDs. Earplugs preserve many individual spectral cues, while earmuffs cancel out most of these cues. We compare methods of combining a simulated, generic HRTF with the simulated, individual Pinna Related Transfer Function. An analytic model of HRTFs, controlled by the azimuth angle, is developed. Respecting the limitations of embedded systems, regarding energy supply and computational power, 14 digital filters are defined. A headphone based listening test is conducted to rate these filters regarding subjective front-back discrimination performance, resulting in better performance with low order filters than with high order filters. We present 4 designs of advanced HPDs which are aimed to improve the sound localization performance. Prototypes are manufactured and evaluated in a subjective listening test with 36 participants, showing that it is possible to improve sound localization of a commercially available active HPD.

Keywords : Binaural filter, Front-back confusion, Hearing protection device, HRTF measurement, HRTF modifications, HRTF simulation, Localization performance, Non-individual HRTF, Subjective listening test.

ABSTRACT

Résumé

De nombreux professionnels sont exposés à des bruits impulsifs ou constants, de très forts niveaux. Pour se prémunir d'une perte auditive, ils portent des protecteurs auditifs. Il s'ensuit une réduction de la localisation des sources sonores, cependant celle-ci est importante, notamment pour des raisons de sécurité. L'objectif de cette étude est de concevoir des systèmes qui concilient la protection auditive, tout en gardant la perception d'espace. Un environnement acoustique virtuel (EAV) de 16 haut-parleurs disposés circulairement et horizontalement est mis en place pour tester les systèmes de protection. La performance de localisation avec des protections acoustiques de types actives et passives, bouchons et casques, disponibles sur le marché, est évaluée avec 40 sujets. Ce test a montré que les bouchons sont à préférer aux casques, selon ce critère de conservation des capacités de localisation. Ces résultats sont corrélés avec les modifications des Head Related Transfer Functions (HRTFs) introduites par les protecteurs auditifs. On montre que les bouchons conservent mieux que les casques les indices spectraux individuels. On compare 2 méthodes pour combiner une HRTF générique et pré-simulée, avec la fonction de transfert individuelle relative à la coque simulée. Un modèle analytique des HRTFs, contrôlé par l'angle d'azimut, est développé. Avec les contraintes imposées, en termes de ressources d'énergie et de puissance de calcul sur des systèmes embarqués, 14 filtres sont définis. Un test d'écoute a permis d'évaluer ces filtres, concernant la discrimination des sons émis devant et derrière l'auditeur, et les filtres d'ordre faible montrent de meilleurs résultats que ceux d'ordre élevé. On propose 4 approches pour des protecteurs auditifs avancés de type casque, qui ont comme but d'améliorer la localisation des sons. Ces prototypes ont été assemblés, et évalués grâce à un test d'écoute avec 36 participants qui montre qu'il est possible d'améliorer la performance de localisation d'un protecteur auditif du commerce.

Mots-clés : Confusion avant-arrière, Filtre binaural, HRTF non-individualisée, Mesure d'HRTF, Modification d'HRTF, Performance de localisation, Protecteur auditif, Simulation d'HRTF, Test sub-

RESUME

jective d'écoute.

Contents

Acknowledgments	3
Abstract	5
Résumé	7
List of tables	17
List of figures	23
Introduction	25
1 Theoretical background & Methods	29
1.1 3D Space	30
1.2 Human auditory system	30
1.2.1 Sound perception	31
1.2.1.1 Loudness	32
1.2.1.2 Noise exposition	33
1.2.2 Sound localization	34
1.2.2.1 Interaural time difference	34
1.2.2.2 Interaural level difference	35
1.2.2.3 Cone of Confusion	37

CONTENTS

1.2.2.4	Spectral cues	38
1.2.2.5	Head-related transfer function	40
1.3	Principles of hearing protections	42
1.3.1	Hearing loss	42
1.3.2	Legal Situation	43
1.3.3	Hearing protection devices	43
1.3.3.1	Fitting method	44
1.3.3.1.1	Earplug	44
1.3.3.1.2	Earmuff	44
1.3.3.2	Attenuation method	45
1.3.3.2.1	Passive attenuation	45
1.3.3.2.2	ANR/ANC	45
1.3.3.2.3	Pass-through	46
1.3.3.2.4	ANR/ANC & pass-through	46
1.3.3.3	Double Protection	46
1.3.4	Issues with HPDs	47
1.3.4.1	Compatibility	47
1.3.4.2	Fitting	48
1.3.4.3	Communication	48
1.3.4.4	Sound localization	49
1.4	Facilities	51
1.4.1	Acquisition of HRTFs	51
1.4.1.1	Acoustic measurement	52
1.4.1.1.1	Anechoic chamber	52
1.4.1.1.2	Setup	53

CONTENTS

1.4.1.1.3	Dummy heads	53
1.4.1.1.4	Post processing	55
1.4.1.2	Numerical simulation	56
1.4.2	Virtual acoustic environment	57
1.4.2.1	Initial setup	58
1.4.2.2	Advanced setup	58
1.4.2.2.1	Reverberation Time	58
1.4.2.2.2	Loudspeaker transfer function	59
1.4.2.2.3	Hardware configuration	59
1.4.2.2.4	Controlling software	61
1.4.2.2.5	Sound source independence	63
1.4.2.2.6	Demonstration software	64
2	Interaction of hearing protection and sound localization	65
2.1	Hearing conditions	66
2.2	Head-related transfer function	67
2.2.1	Measurement setup	67
2.2.2	Obtained HRTFs	67
2.2.3	Changes in the HRTF	68
2.2.3.1	Frequency banks	72
2.2.3.2	HPD induced modifications	73
2.2.3.3	Earplug induced modifications	77
2.3	Localization test	78
2.3.1	Setup	78
2.3.2	Evaluation metadata	79
2.3.2.1	Response Time	80

CONTENTS

2.3.2.2	Head position	82
2.3.3	Localization results	82
2.3.3.1	Precision	85
2.3.3.2	Angular error	86
2.3.3.3	Correct responses	87
2.3.3.4	Correct quadrant	89
2.3.3.5	Confusion rate	91
2.3.3.5.1	Front-back confusion	91
2.3.3.5.2	Left-right confusion	93
2.3.3.5.3	Mixed confusion	93
2.3.3.6	Front-back confusion rate & hearing level	94
2.4	Conclusion	94
3	Investigations on the HRTF and its components	99
3.1	HRTF simulation	100
3.2	Contributions of the outer ear	103
3.3	Recombination of the HRTF	105
3.4	Conclusion	107
4	Spatial audio filters	109
4.1	Parametric Difference-HRTF	110
4.1.1	Front-back differences in HRTF	110
4.1.2	Parameterized filters	111
4.1.2.1	Basic design	113
4.1.2.2	Center frequency and bandwidth	113
4.1.2.3	Gains	115
4.1.2.4	Gain optimization	115

CONTENTS

4.1.2.5	Filter combination	116
4.2	Listening Test	120
4.2.1	Setup	120
4.2.2	Stimuli	120
4.2.2.1	Rendering Filters	123
4.2.2.1.1	High shelf filter	123
4.2.2.1.2	Blauert band filters	124
4.2.2.1.3	Parameterized difference-HRTF	124
4.2.2.1.4	Averaged difference-HRTF	126
4.2.2.1.5	Averaged HRTF	126
4.2.2.1.6	Simulated HRTF	127
4.2.3	Procedure	128
4.2.4	Participants	130
4.2.5	Results	130
4.2.5.1	Response time	131
4.2.5.2	General performance	131
4.2.5.3	Angular error	133
4.2.5.4	Correct responses	134
4.2.5.5	Confusion rate	136
4.2.5.5.1	Front-back confusion	136
4.2.5.5.2	Left-right confusion	139
4.2.5.5.3	Mixed confusion	139
4.2.5.6	Interactions	140
4.2.5.6.1	Angular error & Response time	140
4.2.5.6.2	Front-back confusion rate & Hearing threshold	140

CONTENTS

4.2.5.6.3	Front-back confusion rate & Series	142
4.2.6	Discussion	142
4.3	Conclusion	147
5	HPD prototypes	151
5.1	Design aims	152
5.2	HPD Prototypes	152
5.2.1	Prototype A	153
5.2.2	Prototype B	159
5.2.3	Prototype C	160
5.2.4	Prototype D	163
5.3	Spectral cues	163
5.4	Localization performance	168
5.4.1	Stimuli	169
5.4.2	Setup	169
5.4.3	Procedure	171
5.4.4	Participants	172
5.4.5	Results	172
5.4.5.1	Response time	172
5.4.5.2	Replay	173
5.4.5.3	Preferred sound & Sound imagination	174
5.4.5.4	Confusion matrices	174
5.4.5.5	Localization performance	175
5.4.5.5.1	Correct responses	175
5.4.5.5.2	Front-back confusions	177
5.4.5.5.3	Left-right confusions	178

CONTENTS

5.4.5.5.4	Mixed confusions	179
5.4.6	Discussion	180
5.5	Conclusion	183
Conclusion		185
6.1	Results	185
6.2	Perspectives	187
Bibliography		189
A	Résumé étendu des travaux de thèse	209
A.1	Chapitre 1	209
A.2	Chapitre 2	210
A.3	Chapitre 3	215
A.4	Chapitre 4	218
A.5	Chapitre 5	221
A.6	Conclusion	226
B	Ambisonic encoder test	229
B.1	Stimuli	229
B.2	Experimental procedure	230
B.3	Results	231
B.3.1	Angular error	231
B.3.2	Confusion Matrix	233
B.3.3	Angular error per direction	234
B.3.4	Localization errors	235
B.4	Encoded signals	236
B.5	Conclusion	238

CONTENTS

C	List of acronyms	241
D	List of symbols	245

List of Tables

1.1	Blauert's directional & boosted bands	40
1.2	HPD models	44
1.3	Reverberation time RT60 of anechoic chamber	53
1.4	Reverberation time RT60 of semi-anechoic chamber	59
4.1	Center frequencies and bandwidth of peak filters	114
4.2	RMSE of $\hat{\Delta}$ HRTF	117
5.1	Confusion tables of quadrants and directions of rotation	175
5.2	Detailed front-back confusion rates	179

LIST OF TABLES

List of Figures

1.1	Orientation coordinate system	31
1.2	Definitions of planes	31
1.3	Threshold of hearing & equal loudness	33
1.4	Weighting functions	33
1.5	Illustration ITD and ILD	36
1.6	Calculation and measurement of the ITD	36
1.7	Calculation and measurement of the ILD	38
1.8	Cone of confusion	39
1.9	Directional cues	39
1.10	Use of earplug HPD	45
1.11	Closed loop control of active HPD	46
1.12	Open loop control of active HPD	46
1.13	Level depended gain of active HPD	47
1.14	Floor plan anechoic chamber	52
1.15	Frequency response of speakers	54
1.16	Signal flowchar HRTF measurement	54
1.17	Measurement setup in anechoic chamber	54
1.18	Dummy head Harry34	55
1.19	Dummy head Harry33	55

LIST OF FIGURES

1.20	Dummy head HATS33	55
1.21	Influence of loudspeakers	60
1.22	Setup of listening test in semi anechoic chamber	60
1.23	Sketch of VAE setup	61
1.24	GUI of the controlling software	63
2.1	Hearing protection device P1	66
2.2	Hearing protection device P2	67
2.3	Hearing protection device P3	67
2.4	Hearing protection device P4	67
2.5	HRTFs Harry34 measured	69
2.6	HRTFs Harry33 measured	70
2.7	HRTFs HATS33 measured	71
2.8	HRTFs with custom earplugs	72
2.9	Filter banks for RMSE calculation	73
2.10	RMSE in critical bands	74
2.11	RMSE in Blauert bands	75
2.12	Influence of earplug type on HRTF	77
2.13	Distribution subjects' age	80
2.14	Response time	81
2.15	Response time	81
2.16	Head position versus response time	83
2.17	Confusion matrices localization test.	84
2.19	Signed angular error distribution	87
2.20	Unsigned angular error vs test progress	87
2.21	Correct response rate filters	89

LIST OF FIGURES

2.22	correct response rate trend	90
2.25	Confusion rates	92
2.26	Hering level versus confusion rates	95
2.27	Relation correct response rate and RMSE	97
3.1	Numerical simulation of HRTFs of different dummy heads	101
3.2	HRTF: simulation vs measurement	102
3.3	Numerical simulation of HATS HRTFs	104
3.4	Error function of HRTF composition	106
4.1	Spectral front-back differences	112
4.2	Approximation of difference HRTF	113
4.3	Structure of $\hat{\Delta}$ HRTF.	118
4.4	Output gain $\gamma_o(\phi)$ of $\hat{\Delta}$ HRTF	119
4.5	Parametric spectral front-back difference-HRTF	119
4.6	Error function of $\hat{\Delta}$ HRTF	119
4.7	User interface for listening test	121
4.8	Spectrogram of sound samples	121
4.9	Ambisonic processing chain	122
4.10	Discrete processing chain	123
4.11	Listening test: 3 kHz High shelf	124
4.12	Implementation of Blauert's directional and boosted bands.	125
4.13	Listening test: $\hat{\Delta}$ HRTF based filters	125
4.14	Listening test: $\overline{\Delta}$ HRTF based filters	126
4.17	Distribution participants' age	130
4.18	Response time	132
4.20	Distribution and trend of angular error of both series.	134

LIST OF FIGURES

4.21	Angular error per angle of incidence	135
4.22	Correct response rate	137
4.23	Confusion rates	138
4.24	Unsigned angular error over response time	141
4.25	Subjects' front-back confusion rate over the hearing level.	143
4.26	Front-back confusion rate between series	144
5.1	Sketch of Prototype A	154
5.2	3D model Prototype A	154
5.3	Prototype A	155
5.4	Microphone directivity	156
5.5	Directivity pattern of Prototype A	157
5.6	Summed directivity pattern of Prototype A	158
5.7	Sketch of Prototype B	159
5.8	3D model Prototype B	160
5.9	Prototype B	160
5.10	Sketch of Prototype C	161
5.11	Prototype C	161
5.12	Directivity pattern of Prototype C	162
5.13	Sketch of Prototype D	164
5.14	3D model Prototype D	164
5.15	Prototype D	164
5.16	Hearing condition HC0 to HC5	166
5.18	Cross-sectional view of Prototype D	168
5.19	Spectrogram of sound samples	170
5.20	Graphical User Interface (GUI) for the listening test.	170

LIST OF FIGURES

5.21	Trend of response time along test progress	173
5.22	Box plot with mean values (asterisks) of correct responses for the hearing conditions HC0 to HC5.	176
5.24	Box plot with mean values (asterisks) of front-back confusions for each of the hearing conditions.	178
5.25	Box plot with mean values (asterisks) of left-right confusions for each of the hearing conditions.	179
5.26	Box plot with mean values (asterisks) of mixed confusions for each of the hearing conditions.	180
B.1	Response sheet encoder test	232
B.2	Angular error of ambisonic encoder test	233
B.3	Confusion matrices of ambisonic encoder test	234
B.4	Angular error over tested angle	235
B.6	Front-back confusion rates of ambisonic encoder test	236
B.7	B-Format signals	238

LIST OF FIGURES

Introduction

The auditory system is an important organ for humans to interact with the environment by providing auditory perception of the surroundings. It facilitates participation in life, allowing verbal communication and social integration, and recognizing and localizing sounds and hazards. The auditory system is one of the most skilled, probably the best developed, human sense and in terms of sensory performance it disposes of outstanding specifications. The dynamic range is of factor 10000000 (ten million!), the frequency range spans from 20 Hz up to 20 000 Hz, and the temporal resolution is 2 ms. Not only single sound sources, but also multiple, simultaneous sound sources can be processed with these performance parameters.

In environments with high level noise, it is worth protecting the auditory system from damage. Auditory pain caused by high level noise is a signal of a hazard situation and the most evident, subconscious reaction is to cover the ears with the palms of the hand and trying to escape the situation. In contrast, this prevents the hands from other operations, such as defending or working, though this solution is not useful for professionals. They use Hearing Protection Devices (HPDs), which simultaneously provide protection and readiness for operation. Unfortunately, HPDs have drawbacks, as the reduction of speech intelligibility, face-to-face communications, and sound localization performance. On the one hand HPDs protect well against auditory diseases, but on the other hand, the drawbacks increase the health risks. The user may not perceive correctly or in time hazards and may not respond adequately to dangerous situations. In general, users rank possible hearing loss less severe than face-to-face communications and situational awareness and though even being exposed to high noise level, they do not wear HPDs.

This dilemma is picked up to state the starting point of the following study, which presents principles and concepts of solutions for an acoustically “transparent” hearing protection aiming to reconstruct the auditory environment under an HPD. The concept of the ideal HPD sketches a device

which provides perfect protection, enables face-to-and radio communication, and allows sound source localization with performance equal to natural hearing. Enriching the audio that is presented under the HPD to the user with spatial information allows to step closer to this ideal HPD.

This thesis presents the work that was conducted during a 3.5-year study at the French-German Research Institute of Saint-Louis (ISL). The first chapter focuses on the background knowledge about the human auditory system and HPDs, discusses high level noise and hearing loss. Further, tools and facilitates were developed and are presented which will then be used in the following chapters.

A profound evaluation about HPDs, the HRTF and the sound localization performance is presented in the second chapter. Objective measurements of the natural HRTFs and multiple, non-natural HRTFs of three dummy heads are performed in an anechoic chamber. The non-natural HRTFs are obtained by deploying four different HPDs, including active and passive, earplugs and earmuffs. These measurements reveal a correlation between the type of HPD and the modifications in the HRTF caused by the individual HPDs. Certain HPDs are identified to influence the HRTF less than other HPDs. A subjective localization test is conducted to determine the localization performance by using the same HPD as deployed in the HRTF measurements. The localization performance and the number of front-back errors highly depends on the type of HPD. A direct relation is identified between the objectively measured HPD induced modifications in the HRTF and the subjectively assessed localization performance with the HPDs.

The preceding HRTF measurements are verified in the third chapter by numerical simulations of the natural HRTF of the three dummy heads. The simulation results are confronted with the measurement results, followed by a discussion about their differences and similarities. Further, the HRTF is analyzed more in detail and the contributions by the individual parts of the anatomy are identified. This knowledge is used to propose a method which aims to reduce the extremely time-consuming acquisition of an entire, individual HRTF by recombining the individual and non-individual elements of an HRTF. This chapter concludes on recommendations on earplug HPDs.

The first part of the fourth chapter deals with the development of an analytical model of an HRTF. The motivation for this model is to compress an entire HRTF, while maintaining the front-back cues and representing it by a small number of digital filters. The model is derived from the measured and simulated HRTFs of a dummy head and consists of five 2nd order, Infinite Impulse Response (IIR), peak filters which are optimal superposed. The gains, bandwidths and center frequencies of the peak

filters are controlled by the azimuth angle ϕ , i.e. the model's only parameter. This model of order 10 provides a simple filter which is defined continuously on the entire horizontal plane and shapes the frequency decade between 1 kHz and 10 kHz. This frequency range was identified to contribute most front-back cues and thus is most important for front-back discrimination. The second part of the fourth chapter deals with the reconstruction of front-back cues in sound fields by filtering the sound signals. A large set of filters is tested and evaluated within a headphone based listening test of 45 subjects, split into two groups. The developed HRTF model, a 3 kHz High-Shelf Filter (HSF), equalizing filters based on Blauert's bands, and HRTF filters are comprised in the test set and show that smooth, non-individual filters provide increased front-back discrimination than fine structured, non-individual HRTFs.

The fifth and last chapter presents four HPDs prototypes designs which are intended to increase the sound localization performance. As earmuffs show the worst localization performance we focus on these types of hearing protections for the designs of our prototypes. The prototypes are assembled and evaluated regarding localization performance and usability. To keep them simple in this first stage of development, we mounted them as sensors, not as fully functional protections. The major differences between the designs are the arrangements of the microphone arrays and the way they are attached to the prototype: Six microphones are mounted on the outside the shells of Prototype A, one 1st order ambisonic microphone is mounted to Prototype B, six microphones are mounted on the ballistic helmet of Prototype C, and Prototype D consists of a cavity in the outside of the shells. Directional cues are applied for Prototype A to C by digital filters, while Prototype D uses the principle of directional filtering by the cavity, like the human outer ear. Objective analysis of the provided spectral front-back cues show largely varying results. The prototypes are evaluated with a concluding subjective listening test with 36 participants. The obtained localization performance is in line with the previously obtained spectral front-back cues: HPDs which provide few front-back cues lead to worst localization results. Additionally, it is identified, that simply providing any spectral front-back cues does not automatically increase localization performance. Nevertheless, subjects showed less front-back errors, i.e. better localization performance, with the prototypes than with a commercially available HPD.

Chapter 1

Theoretical background & Methods

Content

1.1	3D Space	30
1.2	Human auditory system	30
1.2.1	Sound perception	31
1.2.2	Sound localization	34
1.3	Principles of hearing protections	42
1.3.1	Hearing loss	42
1.3.2	Legal Situation	43
1.3.3	Hearing protection devices	43
1.3.4	Issues with HPDs	47
1.4	Facilities	51
1.4.1	Acquisition of HRTFs	51
1.4.2	Virtual acoustic environment	57

This first chapter presents the basis for the following work on hearing protections and sound localization. At first, the coordinate system is defined which is used throughout the document. Then, the human auditory system and its main principles for sound perception and sound localization are explained. Both are covered by the wide field of psychoacoustics which deals with the human auditory system and its functionality. Explaining all details would go beyond the scope of this work, though only the relevant aspects are presented. Further, hearing protections are introduced and their need but also issues are given. Finally, tools and setups that are commonly used for different studies are presented.

1.1 3D Space

In the scope of this work the system of coordinate is right-handed and related to the listener or dummy head. The system of coordinate's origin is placed in the center of head (CoH) of the listener or dummy head. The X-axis of the system of coordinate is oriented to the front and the Y-axis is oriented to the left. The Y-axis is in line with the interaural axis. The azimuth angle ϕ increments counterclockwise, starting from the front. ϕ ranges the interval $[0^\circ; 360^\circ]$, such that $\phi = 0^\circ \equiv$ frontal direction, $\phi = 90^\circ \equiv$ left direction, $\phi = 180^\circ \equiv$ back direction, and $\phi = 270^\circ \equiv$ right direction. The elevation angle θ increments from below the listener to above the listener. θ ranges the interval $[-90^\circ; 90^\circ]$, such that $\theta = -90^\circ \equiv$ direction from below, $\theta = 0^\circ \equiv$ direction in the horizontal plane, and $\theta = 90^\circ \equiv$ direction from above. The direction of incidence of a sound is defined by ϕ and θ . These definitions are visualized in Figure 1.1. The normal vector of the horizontal plane is in line with the Z-axis, that of the median plane (also known as median sagittal plane) is in line with the Y-axis, and that of the frontal plane is in line with the X-axis. All three planes intersect in the origin of the coordinate system, while the horizontal plane and the frontal plane intersect in the interaural axis c.f. Figure 1.2.

1.2 Human auditory system

The sound pressure p in Pascal (Pa) describes the variation of the total pressure p_{tot} [Pa] around the atmospheric pressure p_{atm} [Pa] [1]. The temporal variation of p , i.e. caused by vibrating chords, are audible to humans and are considered as sound. In contrast, the temporal changes of p_{atm} , e.g.

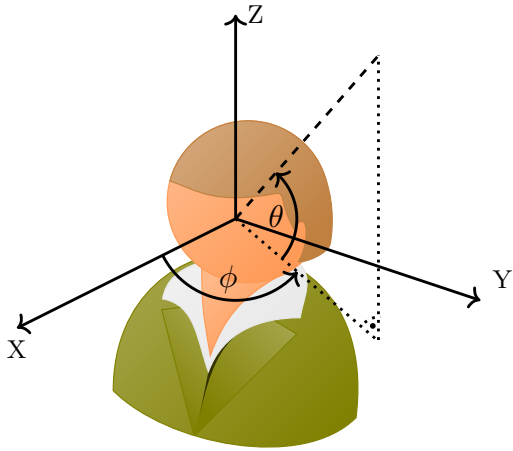


Figure 1.1: Definition of the right-handed coordinate system, placed in the CoH. The direction of incidence (dashed line) is defined by ϕ and θ .

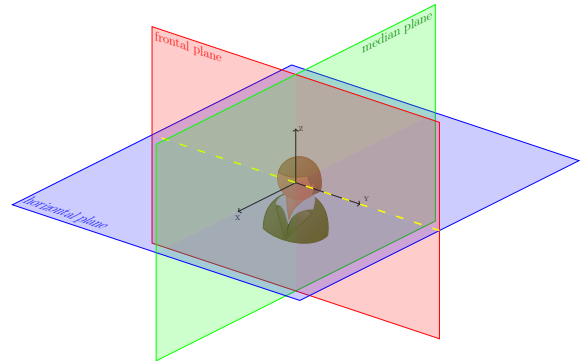


Figure 1.2: Orientation of the horizontal plane (blue plane), the median plane (green plane), and the frontal plane (red plane) relative to the listener. The horizontal and frontal plane intersect on the interaural axis (yellow dotted line), which is in line with the Y-axis.

caused by meteorological phenomena, are too slow and are not audible to humans.

In acoustics it is common practice expressing the sound pressure p by converting the effective sound pressure \tilde{p} to the sound pressure level L_p in dB SPL [2]. This conversion considers the reference sound pressure p_0 of $20 \mu\text{Pa}$, which is close to the threshold of hearing at 1 kHz [3]. The relation between \tilde{p} and L_p is defined in Equation (1.1) [4].

$$L_p = 20 \cdot \log_{10} \left(\frac{\tilde{p}}{p_0} \right) [\text{dB SPL}] \quad (1.1)$$

1.2.1 Sound perception

After passing the outer ear, incoming sound enters the ear canal and reaches the eardrum. From the eardrum it is transmitted via the auditory ossicles to the oval window of the cochlea, exciting the basilar membrane. Hair cells, distributed along the entire basilar membrane, get displaced and convert the excitation into neural signals which are received and evaluated by the auditory cortex [5, 6]. The distance from the oval window to the position of the maximum excitation of the basilar membrane encodes logarithmically the excitation frequency. High frequencies excite the basilar membrane next to the oval window, low frequencies excite the basilar membrane next to the helicotrema [4, 7]. The excitation of the basilar membrane caused by pure tones can be modeled as gammatone filters [8, 9].

The frequency resolution of the auditory system increases with the frequency and varies between 1 Hz and 1.8 Hz [10]. The auditory system processes the incoming auditory spectrum in frequency bands which are much larger than the frequency resolution [11]. These frequency bands, defined by Zwicker and called “critical bands”, split the audible frequency range in 24 frequency bands of varying bandwidths. These frequency bands are of constant bandwidth of $f_b = 100$ Hz for $f \leq 500$ Hz and correspond to one-third octave bands for $f > 500$ Hz [12].

The human auditory system perceives frequencies between 20 Hz and 20 kHz [1]. The absolute threshold of hearing, i.e. the minimum amplitude of a sound to be just noticeable, in free field is shown in Figure 1.3 as it is defined in [13]. The absolute threshold of hearing depends on the frequency while low and high frequencies required a much larger amplitude than mid-range frequencies in order to be perceptible. Humans are most sensitive to frequencies next to the resonance frequency of the ear canal ($\lambda/4$ resonator with a length of 20 mm) of around 4 kHz, c.f. global minimum between 3 kHz and 4 kHz of solid line in Figure 1.3 [4]. The audible sound pressure for frequencies between 50 Hz and 10 kHz ranges from the threshold of hearing of approx. -2 dB SPL to the threshold of pain between 120 dB SPL and 140 dB SPL [4]. Sound pressures above the threshold of pain is perceived but pose health risks. The Loudness Discomfort Level (LDL) ranges from 86 dB HL to 98 dB HL [14], or from 87.4 dB HL to 93.0 dB HL [15], or is equal to 100 dB HL [16]. dB HL denotes the hearing level, indicating the sound pressure level above the subject’s individual threshold of hearing.

1.2.1.1 Loudness

The physiologically perceived loudness L_{phon} of a sound is measured in “phon” and depends on the frequency and the sound pressure of the sound [17]. One phon equals the perceived loudness of a 1 kHz pure tone at 1 dB SPL. This loudness metric allows to compare the perceived loudness of pure tones, considering the frequency dependent sensitivity of the auditory system [18]. Contours of equal loudness were experimentally determined with subjective listening test where the stimuli were presented through headphones [19] or through a loudspeaker [20]. A pure tone under test is perceived equally loud as a 1 kHz pure tone which is located on the same equal loudness contour as the pure tone under test. Tones of different frequencies that are perceived as equally loud do not necessarily have to be presented at equal sound pressure levels, e.g. points (A) and (B) in Figure 1.3. Further, tones of different frequencies and equal sound pressure levels are perceived differently loud, e.g. points

1.2. HUMAN AUDITORY SYSTEM

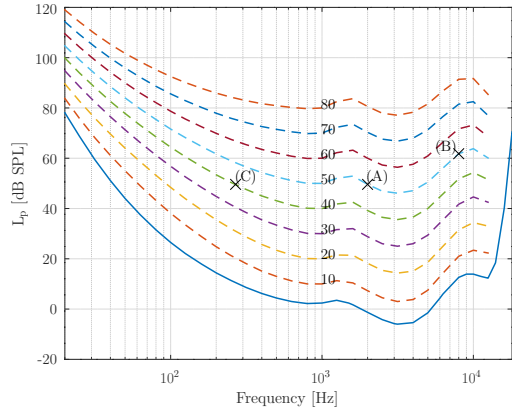


Figure 1.3: Absolute threshold of hearing (solid line) [22]. Contours of equal loudness level from 10 phon to 80 phon (dashed lines) [13]. See text for (A), (B), and (C).

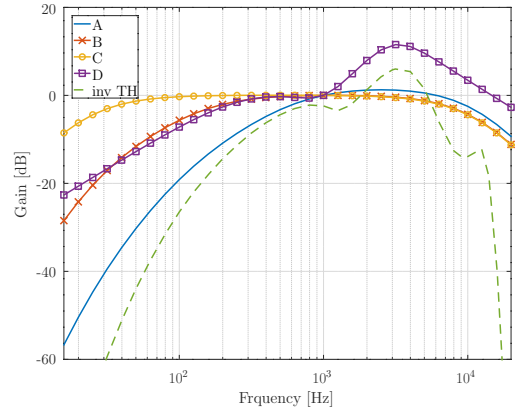


Figure 1.4: A-, B-, C-, and D-weighting functions [21, 23, 24, 25] and inverse threshold of hearing (inv TH) [13] between 16 Hz and 20 kHz.

(A) and (C) in Figure 1.3.

These equal loudness curves are the basic principle for weighting functions, which are used to assess the perceived loudness of broadband signals, such as complex tones or noise signals. The weighting functions are approximations of inverse loudness contours [18]. These weighting functions allow objective measurements to be made that consider the physiological characteristics of the non-linear hearing threshold. The A-weighting curve approximates the 30 phon and 40 phon contours [17, 21] and is used for general assessment of sounds, the B-weighting curve approximations the 70 phon and 80 phon contours [17, 21] and is used for intermediate levels, the C-weighting curve is nearly flat and used for high level sounds [17], while the D-weighting curve takes again into account the ear canal resonance frequency and is used for extreme high level noises, such as aircraft noise [21]. These weighting functions are shown in Figure 1.4.

1.2.1.2 Noise exposition

The equivalent continuous sound level $L_{eq,t}$ is used to attribute a constant sound pressure level to a noise with a time varying sound pressure level in the time interval t . The energy of the original noise with time varying level is equal the energy of the same noise but at constant level $L_{eq,t}$. $L_{eq,t}$ allows to compare multiple noises of different levels and duration. Integrating L_{eq,t_i} for each noise event i of duration t_i over the standard daily working time of 8 h results in the daily noise exposition level

$L_{ex,8h}$, c.f. Equation (1.2). This expression allows to monitor the exposition of employees to noise during working time and is calculated on the A-weighted $L_{eq,t}$ [26].

$$L_{ex,8h} = 10 \cdot \log_{10} \left(\frac{1}{8h} \sum_{i=1}^n \left(10^{0.1 \cdot L_{eq,t_i}} \cdot t_i \right) \right) [\text{dB (A)}] \quad (1.2)$$

1.2.2 Sound localization

Binaural and monaural cues are used to determine the location of a sound source, to focus on a single sound source when multiple sources are simultaneously present, called “cocktail party effect”, or to unmask sounds or single frequencies [27, 28]. Binaural cues, such as the Interaural Time Difference (ITD) and Interaural Level Difference (ILD), are determined by combining the acoustic information received by both ears. The incoming sound at both ears is processed in a stereo manner. In 1907, John William Strutt, also known as Lord Rayleigh, introduced the Duplex theory [29]. It describes the principles of sound localization in the horizontal based on the ITD and ILD. For the determination of monaural cues, such as the loudness and spectral cues, the acoustic information received by both ear is separately processed. Even though, the spectral cues are determined separately for both ears, it is assumed that also the spectral difference between both ears is evaluated for the sound localization [30].

1.2.2.1 Interaural time difference

The head is approximated by a sphere with a diameter of 0.21 m [31, 32] and the depth of the ear canals is neglected. This assumes that the auditory system is positioned at the entrance of the ear canal. Hence, we assume the auditory systems to be placed on a sphere of $\varnothing 0.21$ m at diametric opposing positions.

For sound sources which are positioned off the median plane this spacing leads to two different path lengths between the ipsilateral and contralateral sound path, c.f. Figure 1.5. With a constant speed of sound on both paths, the path length difference causes different traveling times on the paths and hence different Time Of Arrivals (TOAs) at both ears. The resulting difference in the TOAs is called ITD, c.f. right part of Figure 1.5.

The ITD is defined for frequencies below 800 Hz, i.e. for wavelengths larger than 0.43 m [33]. For

wavelength smaller than twice the head diameter, the phase shift between both ears exceeds 360° , avoiding a reliable determination of the ITD [34]. Assuming the auditory systems to be placed on the interaural axis without head in between the ITD can be expressed as a function of the azimuth angle ϕ , c.f. Equation (1.3). Using the approximation of the head by a sphere with the auditory systems placed on diametrical opposite positions, we obtain a more detailed function, see Equation (1.4) [32]. The speed of sound c equals 343.23 m s^{-1} at a temperature of 20° and the head radius r equals 0.105 m . For Equation (1.4) the azimuth angle ϕ has to be converted into radians ranging between $\mp 90^\circ$. Both functions are visualized in Figure 1.6.

$$ITD(\phi) = \frac{2r}{c} \cdot \sin(\phi) \quad (1.3) \qquad ITD(\phi) = \frac{r}{c} \cdot (\phi + \sin(\phi)) \quad (1.4)$$

Sounds that are positioned on the median plane ($\phi \in \{0^\circ, 180^\circ\}$) result in angles of incidence that are orthogonal to the interaural axis. The ipsilateral and contralateral path lengths are equal and so the ITD is 0 s . The maxima of the ITD are reached for sounds which are positioned on the interaural axis, maximizing the interaural path difference.

The ITD was measured in an anechoic chamber, using a dummy head on a turntable. A loudspeaker emitted pulsed tone sweep from 62.5 Hz to 16 kHz . The signal levels of the dummy head's left and right built-in microphones were recorded.¹ The obtained ITD is shown in Figure 1.6. The absolute maxima of the ITD are obtained at lateral positions and equal 0.61 ms (simplified approximation, Equation (1.3)), 0.79 ms (advanced approximation, Equation (1.4)), and 0.81 ms (measurement). Apart of measurement uncertainties, the measured ITD match the advanced ITD on the entire horizontal plane, while the measured ITD and simplified ITD deviate at lateral positions. The advantage of considering the circular part of the contralateral sound path around the head, c.f. Figure 1.5, is mainly seen at lateral positions. The minimum audible ITD of $10 \mu\text{s}$ corresponds to an minimum audible angle of 0.94° [35, 36, 37].

1.2.2.2 Interaural level difference

The head is an obstacle that interacts with the surrounding acoustic field, causing a frequency-dependent acoustic shadow on the contralateral side. Nevertheless, sound arrives at the contralateral ear due to diffraction of the acoustic waves around the head, illustrated by the circular part of the contralateral path in the left part of Figure 1.5. The head shadow, the diffraction, and the increased

¹For more details on the anechoic chamber and the dummy head see Section 1.4.1.

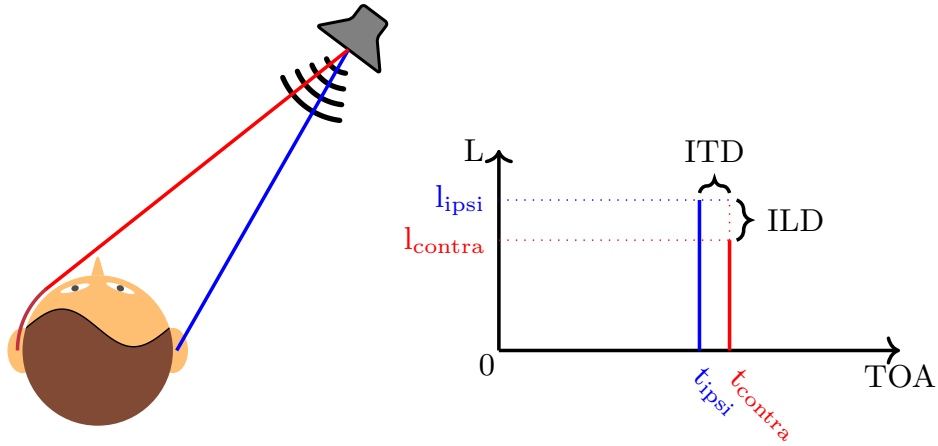


Figure 1.5: Left: Ipsilateral (blue) and contralateral (red) sound paths for a sound source (loudspeaker) with an arbitrary positioning relative to the listener. Right: Sound level (L) over Time Of Arrival (TOA). The sound is emitted by the source at time 0 at an arbitrary level. It arrives at the ipsilateral, resp. contralateral ear at time t_{ipsi} , resp. t_{contra} with level l_{ipsi} , resp. l_{contra} . The difference in the TOA, resp. in the sound pressure level between the left and right ear is marked by ITD, resp. ILD.

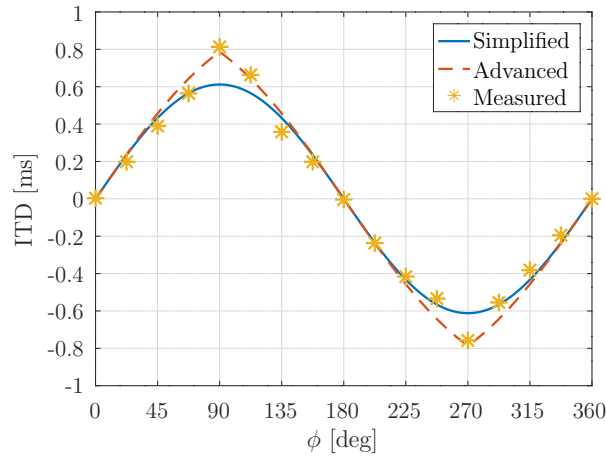


Figure 1.6: The calculated ITD according to Equation (1.3) (“Simplified”) and Equation (1.4) (“Advanced”) and the horizontally measured ITD (“Measured”).

path length result in less acoustic energy arriving at the contralateral ear than at the ipsilateral ear. This difference in energy results in a difference in the sound pressure, which is called ILD, see right part of Figure 1.5.

The ILD is evaluated for frequencies above 1.5 kHz [4, 34], i.e. for wavelengths shorter than 0.23 m. The dimensions of the head are too small to form an obstacle that interacts with frequencies below 1.5 kHz. Particularly for frequencies below 850 Hz, i.e. wavelengths above twice the head diameter, the ILD converges to 0 dB and is no longer determined [34]. Regardless of the frequency, no ILD is introduced for sounds positioned on the median plane. Like for the ITD, these positions result in equal sound path for both ears. The maximum ILD is expected to be introduced by sounds which are positioned at lateral positions. Regarding Figure 1.7, we obtain that the ILD reaches its maximum values just slightly off the interaural axis directions. The decrease of the ILD at lateral positions is due to an “acoustic bright spot”. The diffraction paths around the head are constructively interfering at the diametric opposing position of the direction of incidence ($\phi + 180^\circ$). For directions of incidence which are on the interaural axis, this point of constructive interference matches the position of the contralateral ear. The sound level at the contralateral ear increases and hence the ILD decreases.

Figure 1.7 shows the ILD for nine octave bands between 62.5 Hz and 16 kHz at 16 angles of incidence in the horizontal plane. The simulated ILD in Figure 1.7a is based on a simplified head model, i.e. a sphere with a diameter of 0.21 m. The measured ILD in Figure 1.7b was done with a dummy head in an anechoic chamber.² Well visible are the extreme values that are located off the interaural axis at $\phi \in \{67.5^\circ, 112.5^\circ, 157.5^\circ, 202.5^\circ\}$ and the “acoustic bright spot” effect on the interaural axis, and below 500 Hz only little ILD is introduced.

1.2.2.3 Cone of Confusion

Both the ITD and the ILD are symmetric non-bijective functions, e.g. it is not possible to map the ITD or ILD to one single angle of incidence ϕ . Sound localization based on the Duplex theory leads to ambiguities in the 2D space of the horizontal plane. The ITD and ILD result in equal values for pairs of sound sources which are arranged symmetrically around the interaural axis. The listener cannot discriminate between frontal and back sound incidence which leads to front-back confusions [38]. Extending the space to three dimensions, not only ϕ but also θ has to be determined for unique

²For more details on the anechoic chamber and the dummy head see Section 1.4.1.

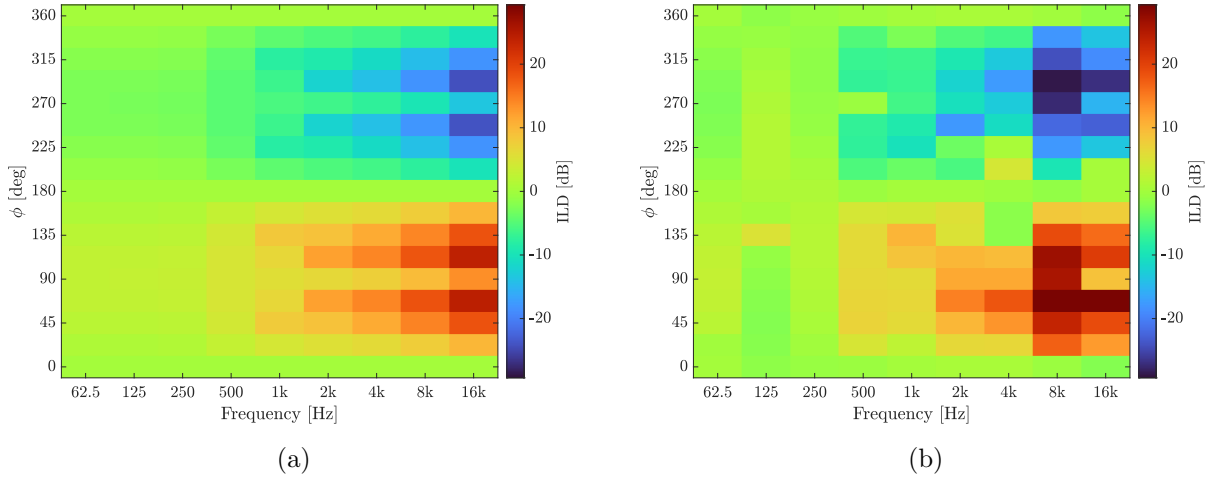


Figure 1.7: Color coded ILD as numerically simulated on a sphere of diameter 0.21 m (Figure 1.7a) and as measured on a dummy head in an anechoic chamber (Figure 1.7b).

direction identification. Using the ITD and ILD for sound localization in the horizontal 2D space leads to a pair of possible angles of incidence, i.e. front-back confusions. Sound localization in the entire 3D space using the ITD and ILD extends this set of front-back confusions to approximately the shape of a cone, which is called cone of confusion. A cone of confusion defines a set of angles of incidence which result in equal values for the binaural cues. An arbitrary cone of confusion is illustrated in Figure 1.8. The cone of confusion is the set of points which are positioned on a symmetric cone, with its apex located in the CoH and its base area orientated perpendicular to the interaural axis, pointing either to the left or right side of the listener [39].

1.2.2.4 Spectral cues

To resolve the cone of confusion and identify a single direction of incidence, the spatial information encoded in the spectral cues is evaluated. The spectral cues contain information about the front-back positioning and the elevation of a sound source. The combination of the ILD, the ITD, and the spectral cues allows a proper determination of ϕ and θ and hence the direction of incidence [38].

Spectral cues are introduced when the incoming sound is reflected, scattered, and diffracted at the listener's shoulder, head, and outer ear. Each frequency component of the sound passes the geometry shoulder-head-outer ear on multiple different sound paths, which depend on the direction of incidence. All these individual sound paths sum up at the eardrum, leading to interference and

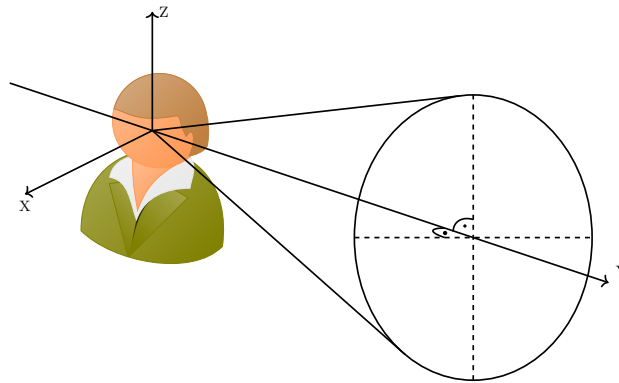


Figure 1.8: Illustration of an arbitrary cone of confusion around the right ear. Evaluating only the ITD and ILD for sound localization in 3D space results in an under-determined set of angles of incidence which are located on a cone.

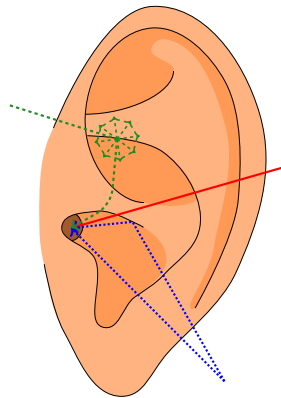


Figure 1.9: Illustration of three arbitrary sound paths from different directions of incidence. Direct path (**solid line**), interference of direct path and reflected path (**dotted line**), and scattered and diffracted sound path (dashed line).

delays. Figure 1.9 illustrates three possible paths for various directions of incidence. Thus, before the incoming sound arrives at the eardrum, spectral cues are added to it that encode the direction of incidence of the sound. Applying these spectral cues to the incoming sound corresponds to filtering the incoming sound with a corresponding filter. The frequency responses of these filters are defined by the azimuth and elevation dependent spectral cues. Each filter belongs to a single shoulder-head-outer ear geometry, azimuth angle and elevation angle. The entire set of filters of a single shoulder-head-outer ear geometry is called Head-Related Transfer Function (HRTF). The HRTF is dependent on the frequency, the azimuth angle, and the elevation angle.

It has been carried out that amplifying certain frequency ranges makes the auditory system localize

1.2. HUMAN AUDITORY SYSTEM

	dir		boost	
	90 %	95 %	90 %	95 %
FR	[280 Hz; 570 Hz] [2.91 kHz; 5.65 kHz]	[2.91 kHz; 3.64 kHz]	[130 Hz; 690 Hz] [1.86 kHz; 7.03 kHz] [13.35 kHz; 16.42 kHz]	[150 Hz; 570 Hz] [1.85 kHz; 2.84 kHz] [3.64 kHz; 5.77 kHz]
RE	[730 Hz; 1.74 kHz] [9.59 kHz; 14.74 kHz]	[730 Hz; 1.73 kHz]	[730 Hz; 1.74 kHz] [7.44 kHz; 12.79 kHz]	[730 Hz; 1.72 kHz] [7.50 kHz; 11.54 kHz]

Table 1.1: Directional (“dir”) and boosted (“boost”) bands of 90 % and 95 % confidence for frontal (“Fr”) and back (“Re”) sound incidence. Data extracted from [40].

the sound in the front, back, or above [40, 41, 42, 43, 44]. Even though varying definitions of such sets of frequency bands have been proposed, Blauert’s bands overlap best with those proposed by other authors [45]. Blauert’s directional bands were determined by evaluating the subjectively perceived direction of one-third octave sounds. Blauert has determined the boosted bands by measuring the difference in the spectral Sound Pressure Level (SPL) between frontal and back sounds [40]. Table 1.1 shows the definitions of these bands as they were extracted from [40].

1.2.2.5 Head-related transfer function

The anatomy of the shoulder, head and outer ear is individual for each listener, so is the HRTF [46]. During life listeners adapt to their own individual HRTF. HRTFs are not perfectly symmetric between the left and right ear [30], in particular for frequencies above 5 kHz the interaural asymmetry reaches up to 20 dB [47, 48]. Further, the HRTF alters from infancy to adulthood [49]. This is a gradually ongoing process, such that the person can constantly adapt to the changing HRTF. An instantaneously change in the HRTF, by e.g. earplugs, ear molds, or hearing protections, results in a poor sound localization performance [50, 51]. At least with ear molds, subjects adapt to the modified HRTF during a six-week training period and the initial sound localization performance is restored [51].

HRTFs are used for binaural recording and binaural rendering. Binaural recordings are achieved with an acoustic dummy head or with humans wearing earplug microphones. The signal of interest is filtered by the HRTF before it is recorded by the microphones [52]. HRTFs are used to create headphone based Virtual Acoustic Environments (VAEs) where sounds are virtually positioned in space around the listener and binaural signals induce an immense realistic listening impression [53, 54]. The

choice of the binaural filter is highly crucial as the binaural filter controls the front-back and elevation perception of the sound by the listener. The main issue of HRTFs is their individuality and that in listening tests or localization tasks listeners prefer their own HRTF over the foreign, i.e. non-individual HRTFs [55]. To decrease the number of front-back and up-down confusions and increase the good listening experience the listener's individual HRTF should be preferred over non-individual HRTF [56]. Miniature microphones at the entrance of the user's ear canal allow individual binaural recordings [47, 57]. In the case of binaural renderings, the listener's HRTF is measured in advance and applied during the rendering process to the source signal. As it is rarely feasible to consider the individual HRTF of any listener, often non-individual, generic HRTFs are used and listeners have to adapt to the non-individual HRTF. An adaption process helps to reduce the front-back confusion rate by 10 % [58]. The adaption to non-individual HRTFs in a VAE does not necessary require visual feedback but does not lead to same localization performance as obtained with individual HRTFs (17 % front-back confusions with non-individual HRTFs versus 12 % front-back confusion with individual HRTFs) [59].

Binaural filters are often based on HRTFs and different techniques are applied to handle, represent, and simplify them [60]. Individual HRTFs require appropriate facilities to record them in anechoic chambers [61, 62, 63, 64] or to simulate them numerically based on 3D models [65, 66, 67]. This implies great effort, but subjects perceive the sounds filtered with their individual HRTF, like they listen the sounds in daily life with their natural hearing. Different approaches were tested to reduce the effort of HRTF measurement. From a set of non-individual HRTFs, the best matching HRTF was selected by performing search tasks based on localization tests [68] or comparison of anthropometric data [55]. Individualization techniques aim to tune non-individual HRTF to match best for a listener. Correlations between the notches and peaks of an HRTF and the anthropometric data were used to individually fit non-individual HRTFs to subjects [69, 70, 71]. Despite all these efforts, best localization performance is still obtained using individual HRTFs [72, 56].

Approximating an original HRTF by a less complex representation is done to gain simplicity on the HRTF but still obtain good localization performance. Principle component analysis were used to identify a set of basis functions of different HRTFs [73]. Kistler and Wightman showed that the first five basis functions are necessary to obtain approximately equal localization performance as obtained with the original HRTF [74]. Nowak et al. combined 1st order Low-Shelf Filter (LSF) and High-Shelf Filter (HSF) and 2nd order peak filters to approximate HRTFs [75]. An error of 2 dB to the original

HRTF was not exceeded when using at least 12 peak filters.

Further simplification of HRTFs is achieved by abstraction. A certain number of characteristics is extracted from a set of HRTF and represented by low order filters or frequency bands. Frank et al. examined the spectral difference between frontal and back sound incidence based on the HRTF of a *Neumann* “KU100” dummy head. They approximated the spectral front-back difference by a 3 kHz HSF [76]. Different authors identified that certain frequency bands are boosted, resp. damped when the sound is originating from the front, resp. back [40, 42, 43, 44, 77]. Conversely, filters which are based on these frequency bands make subjects perceive sounds in the front or back hemisphere.

1.3 Principles of hearing protections

1.3.1 Hearing loss

The threshold of hearing increases over age [78]. This effect is smaller for females (increase by 45.4 dB between the ages of 19 years and 80 years at 4 kHz) than for males (increase by 55.9 dB between the ages of 19 years and 80 years at 4 kHz) and less for low frequencies (increase by 24.6 dB between the ages of 19 years and 80 years at 500 Hz) than for high frequencies (increase by 65.1 dB between the ages of 19 years and 80 years at 6 kHz) [79]. The increase of the threshold of hearing due to maturing is little for very low frequencies (increase weaker than 10 dB between the ages of 25 years and 61 years below 160 Hz) [80] and strong for very high frequencies (increase by up to 44 dB between the ages of 29 years and 50 years at 14 kHz) [81]. The annual rate of hearing loss, i.e. the increase of the threshold of hearing, increases over age for low frequencies (0.4 dB per annum below the age of 59 years vs. 2.0 dB per annum above the age of 80 years at 500 Hz) and is approximately constant over age for high frequencies (1 dB per annum between the ages of 59 years and 80 years at 4 kHz) [82].

Exposure to high level noise above 85 dB (A) induces auditory and non-auditory health damages. The temporal or permanent increase of the threshold of hearing is the main auditory effect, whereas non-auditory effects include cardiovascular diseases, increased cortisol, stress, and discomfort, reduced cognitive performance, and sleep disturbances [83, 84]. Over 90% of all hearing losses are due to sensorineural hearing loss, i.e. damage to the cochlea and auditory nerve [85]. The rate of hearing loss is increased by long exposition times, wide band noise and high sound pressure levels [86]. The shift in hearing threshold caused by high level noise can be limited by using appropriate Hearing Protection

1.3. PRINCIPLES OF HEARING PROTECTIONS

Devices (HPDs). Further it can be recovered to some extent, by recreation intervals of sufficient length between consecutive noise events [87, 83].

1.3.2 Legal Situation

The use of HPDs is a legal obligation for professionals being exposed to high level noise. Professions with endangering hearing, e.g. road builders, track and airfield workers, machine operators, musicians, are exposed to noise level up to 119 dB(A) [88, 87, 89]. In order to protect them from hearing loss, German legislator³ defines the lower and upper action values for the A-weighted exposure level $L_{ex,8h}$, resp. the C-weighted sound pressure peak level $L_{pC,peak}$, at 80 dB (A) and at 85 dB (A), resp. at 135 dB (C) and at 137 dB (C), see Section 3, Article 6, Paragraph 1 and 2 of [91]. The lower action value marks the limit, where the employer has to provide appropriate hearing protection to the employee. The upper action value marks the limit from which an appropriate hearing protection has to be used. The levels of $L_{ex,8h} = 87$ dB (A) and $L_{pC,peak} = 140$ dB (C) are not to be exceeded under protection [91]. People are also exposed to high level noises outside working hours, e.g. spectators in soccer stadiums ($L_{p,peak} = 105$ dB(A)) or pedestrians next to an urban construction site ($L_{p,peak} = 95$ dB(A)) [92, 93]. Hence to avoid hearing loss due to high noise level the use of HPD is essential for a large variety of people.

1.3.3 Hearing protection devices

To respect legislation and prevent hearing loss, different types of HPDs are available which meet most requirements for the varying fields of applications. Table 1.2 shows different models of HPDs, grouped by their method of fitting as proposed by Berger and Casali in [94]. For each work environment, the appropriate model can be selected depending on the present noise level, the required attenuation level and the compatibility with additional safety equipment. HPDs can be differentiated according to their method of fitting, c.f. Table 1.2, but also to their attenuation method. In the following the focus is set on these two main characteristics.

³Directive 2003/10/EC of the European Community [90] holds for all Member States of the European Union which has to be translated into national legislation.





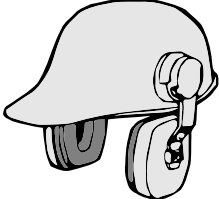
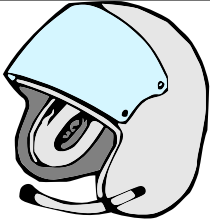
Earplugs		
Premolded	Formable	Custom molded
		
Earmuffs		Helmets
Standalone	Attached	
		

Table 1.2: Common types of HPDs, grouped by their method of fitting. Image adapted from [94].

1.3.3.1 Fitting method

The fitting method of an HPD describes which parts of the head are covered by the HPD and how the device is worn. Among the three groups, shown in Table 1.2, helmets are the least used HPDs. For a wide range of users, first choices are either earplug, also known as in-ear, systems or earmuff, also known as on-ear, systems.

1.3.3.1.1 Earplug Earplug HPDs are inserted into the entrance of the ear canal and partially in the ear canal, illustrated on the left of Figure 1.10. Non-customized earplugs, including premolded models and formable models, and customized earplugs are available, c.f. Table 1.2 category “Earplugs”. The geometry of non-customized earplugs is defined by the manufacturer, while customized earplugs are designed to fit the anatomy of the user’s concha and ear canal entrance.

1.3.3.1.2 Earmuff Earmuff HPDs entirely enclose the outer ear of the user, c.f. right part of Figure 1.10. These systems consist of non-customized cups. Earmuff HPDs are easily interchangeable between users but require a special design to be worn in combination with additional safety equipment, e.g. helmets.

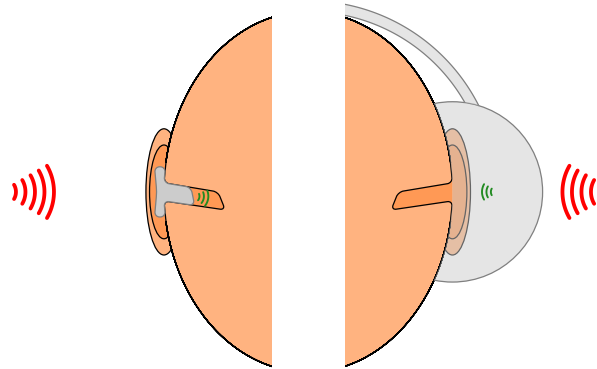


Figure 1.10: Positioning of earplug (left) and earmuff (right) HPDs.

1.3.3.2 Attenuation method

HPD achieve the attenuation either by passive or active technologies. Concerning the active attenuation we distinguished between three different designs.

1.3.3.2.1 Passive attenuation The vibroacoustic characteristics of the material of the HPDs define the attenuation factor of the HPD, hence all HPDs provide passive attenuation. Putting them on or inserting them establishes a sound barrier that attenuates the incoming noise before it reaches the eardrum. The attenuation of passive HPDs is in general fixed, i.e. independent of the surrounding noise level, and of 35 dB. The maximum possible attenuation is limited to 35 dB to 60 dB due to frequency dependent bone-conducted noise [95, 94, 96, 97]. Particular for earplugs, there are designs that allow level dependent attention. These passive non-linear earplugs consist of an opening with a special geometry through which impulse noise is attenuated while sound at non-dangerous levels is allowed to pass through unchanged. In addition to the passive attenuation, certain HPD models incorporate active attenuation techniques. These HPD models are then referred to as active HPD.

1.3.3.2.2 ANR/ANC Equipped with a microphone and loudspeaker, both towards the inside, these systems are based on an Active Noise Reduction (ANR) or Active Noise Control (ANC) technique. The residual noise under the HPD is recorded with the microphone and emitted with opposite phase through the loudspeaker. This re-injected signal is also called anti noise. The residual noise is superposed with the anti noise, leading to destructive interference of both signals. This technique works well at low frequencies with wavelengths much larger than the head dimensions. The control loop of the system

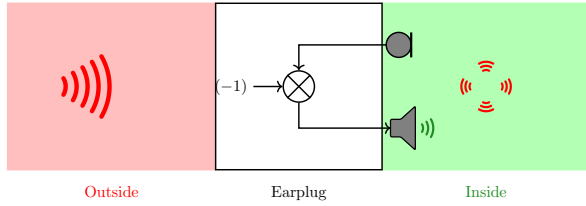


Figure 1.11: Closed loop of an ANR/ANC active HPD: an internal microphone records the residual noise (red). Re-injected with inverted magnitude through the internal loudspeaker superposes the anti noise (green) with the residual noise (red), leading to destructive interference.

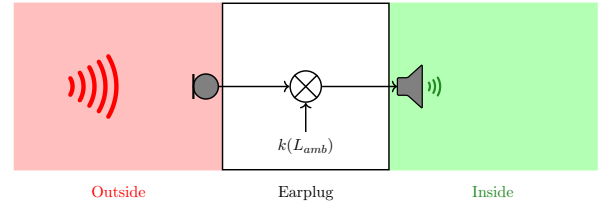


Figure 1.12: Open loop of a pass-through active HPD: an external microphone records the high level ambient noise. The gain $k(L_{amb})$ is applied to the recorded signal and the resulting signal is emitted through an internal loudspeaker at a pleasant level (green).

with the feedback is illustrated in Figure 1.11. The advantage of active HPD over passive HPDs is increased attenuation in the low frequencies. Often, they are not sufficiently attenuated by the vibroacoustic characteristics of passive HPDs.

1.3.3.2.3 Pass-through Equipped with an external microphone and an internal loudspeaker, these systems have a level-dependent dynamic. The surrounding noise is recorded, multiplied with the gain (here called k), and emitted under the HPD at a non-dangerous, pleasant sound level. The gain k depends on the ambient sound pressure level L_{amb} . Weak sounds are amplified, mid-volume sounds are kept unchanged, and high level sounds are attenuated. Figure 1.13 illustrates an exemplary trend of $k(L_{amb})$. The maximum possible attention is limited by the passive attention characteristics of the material of the HPD. The system's control loop with the open loop is illustrated in Figure 1.12.

1.3.3.2.4 ANR/ANC & pass-through Equipped with one external microphone, one internal microphone, and one internal loudspeaker these systems combine the techniques of ANR/ANC (Paragraph 1.3.3.2.2) and pass-through (Paragraph 1.3.3.2.3).

1.3.3.3 Double Protection

The peaks of impulse noise exceed 140 dB in certain situations. This level is equals to the maximum daily noise exposure [98]. This means that the user has already more than his daily noise dose within one noise event. In such situations a single protection might not provide sufficient attention and double protection is required. Earplugs are inserted in the ear canal and earmuffs are additionally covering

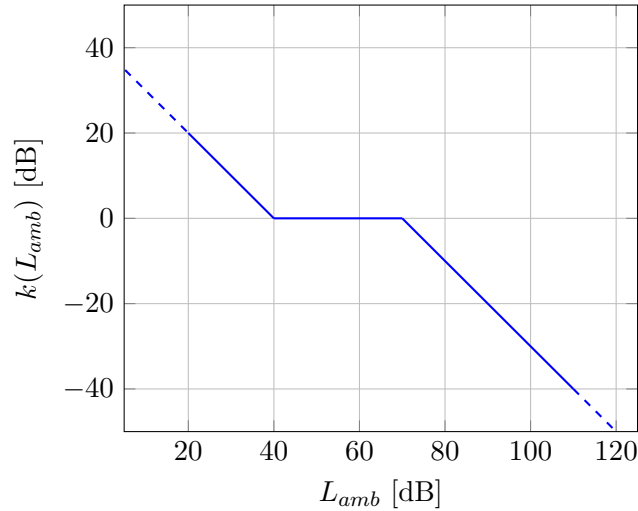


Figure 1.13: The gain $k(L_{amb})$ of a pass-through active HPD depends on the ambient sound level L_{amb} . The pleasant sound level under the HPD can be achieved by amplifying low level sounds ($k > 0$ dB) and attenuating high level sounds ($k < 0$ dB). Here $k(L_{amb})$ is chosen to be entirely monotonous, linearly decreasing for $L_{amb} < 40$ dB and $L_{amb} > 70$ dB, and constant for $L_{amb} \in [40 \text{ dB}; 70 \text{ dB}]$.

the outer ear. The frequency dependent attenuation of double protection is lower than the sum of the attenuation values of the involved HPDs [99, 100]. Double protection shows a much higher attenuation in the lower frequencies than single protection. With increasing frequency this advantage decays and for frequencies from 2 kHz on the attenuation is limited by skull bone conduction [101, 102]. It has been shown that ANR earmuffs lose their benefits when using them in double protection, whereas the attenuation of ANR earmuffs is enhanced [103].

1.3.4 Issues with HPDs

Despite the advantage of HPDs to prevent the user from hearing disorder, there are several drawbacks and problems when using HPDs.

1.3.4.1 Compatibility

Not all models of HPDs can be worn in combination with further personal protective equipment. Especially earmuff HPDs require an appropriate design such that they can be worn simultaneously with hard hats or tactical, ballistic helmets. Solutions of earmuff HPDs which can be mounted to helmets or worn under helmets are illustrated in the lower row of Table 1.2.

1.3.4.2 Fitting

The user must put on or insert correctly the earmuff or earplug HPD. Not perfectly fitting HPDs lead to acoustics leaks and they do not provide safety. The attenuation level of earplugs is reduced by up to 15 dB due to acoustics leaks [94, 97]. Training the insertion of earplugs increases the noise attenuation between 3.6 dB at 2 kHz and 7.6 dB at 8 kHz [104]. One-to-one instructor-trainee sessions, rather than non-personal instructions, such as video or written instructions, for training the correct use of earplugs increases the attenuation by up to 12 dB [105]. For non-experienced users it is necessary to attend training to achieve the attenuation levels specified by the manufacturer of the HPD. Without training, the achieved attenuation levels are up to 16 dB below the specifications [106]. Here solutions have been proposed where active HPDs measure the transfer function of the system {internal loudspeaker – enclosed volume in the ear canal – internal microphone}. Evaluating this transfer function allows to conclude whether the earplug is inserted correctly or not. The HPD informs acoustically the user when it detected acoustic leaks [107].

1.3.4.3 Communication

It has been reported that 22 out of 31 workers have difficulties in either the communication with colleagues or the ability to monitor their environment when wearing HPDs [108]. This subjective impression is supported by a Hearing in Noise Test (HINT) with passive HPDs, which shows a decrease of the speech intelligibility between 15.6 % and 31.4 % compared to the open ear condition [109, 110]. Contrarily, active HPDs boost the speech intelligibility in noisy environments up to 19 % [110]. In situation with moderate background noise ($L_{noise} \in [60 \text{ dB}; 80 \text{ dB}]$), universal foam earplugs reduce the speech intelligibility more (reduction by 28 %) than universal flange earplugs (reduction by 21 %) [111]. Sever hearing loss ($< 41 \text{ dB HL}$) reduces the speech intelligibility by 21.6 % (earplugs) and 55.6 % (earmuffs) while slight hearing loss ($\geq 25 \text{ dB HL}$) reduces the speech intelligibility by 5.6 % (earplugs) and 6.8 % (earmuffs) [109].

Smalt et al. identified with a Modified Rhyme Test (MRT) that the level of the background noise has a much higher effect on speech intelligibility than the choice of the HPD [112]. Compared with the open ear condition, HPDs are only beneficial in environments with negative Signal-to-Noise Ratio (SNR) (increase of intelligibility by up to 7 %) and disadvantageously for positive SNR (reduction of intelligibility by up to 8 %) [113]. Contrary results were obtained by Tufts et al. with passive

1.3. PRINCIPLES OF HEARING PROTECTIONS

earplugs. They identified that the speech intelligibility converges to zero with decreasing SNR, while for $\text{SNR} < -15$ dB no useful speech information is perceived by the user [111].

Initial consonants are correctly perceived in 77 % (earplug) and between 65 % and 82.8 % (earmuff), whereas final consonants are correctly perceived in 80 % (earplug) and between 70 % and 72 % (earmuff) [114, 115]. When using earmuffs for radio or face-to-face communication, the voice of a native speaker leads to an higher MRT score (79.6 %) than the voice of a non-native speaker (MRT score: 75.2 %) [115].

The directions of incidence of the noise and the speech are of importance for the intelligibility. Least intelligibility is obtained (22 % words correctly identified) when the noise and the speech are both originating from the front. Intelligibility increases when the noise is placed lateral and the speech frontal (42 % words correctly identified) and increases further when the noise and speech are of opposing directions (53 % words correctly identified). The effect is the least, resp. most noticeable with active HPDs and their gain set to minimum (increase by 30 % for opposing directions over same directions), resp. passive HPDs (increase by a factor of 3 for opposing directions over same directions) [116]. This is explained by the cocktail party effect. The spatial separation of multiple sound sources present at the same time allows the auditory system to focus on one source and extract its information [27].

1.3.4.4 Sound localization

The loss of sound source localization accuracy by the user and the increased time required to identify the localization of a sound source are often mentioned in the context of HPDs [117, 50, 118, 119, 120, 121]. HPDs lead to an increase of the localization error from 15° to 50° [117], from 8° to 30° [119], or from 15° to 33° [122]. Further, HPDs increase the number of front-back confusions from 3 % to 24 % [117], from 2 % to 17 % [50], or from 4.5 % to 17.4 % [119]. The azimuth error depends on the tested direction and reaches its maximum at back median positions [123].

Subjects decide less frequent for the correct direction of incidence with earmuffs (at most 49.3 % [120], 57 % [124], 40 % [50], 66 % [119]) than earplugs (at least 54.8 % [120], 63 % [124], 59 % [50], 82 % [119]). A similar trend has been obtained with devices for music listening. Environmental sounds are less reliable perceived with earmuffs (reduction of correct rate by 32 % compared to open ear condition) than with earplugs (reduction of correct rate by 7 % compared to open ear condition) [125].

1.3. PRINCIPLES OF HEARING PROTECTIONS

With head movements allowed and either using earplugs or earmuffs, subjects immediately turn their heads in the correct azimuth direction. Concerning the elevation direction, subjects initially turn in the opposite direction, but correct themselves and then turn in the correct direction [118, 121].

Zimpfer et al. reported that passive, non-linear earplugs lead to slightly more correctly perceived sound directions (64 %) than active earplugs (53 %) [126] whereas Brown did not notice any difference in the Root Mean Square Error (RMSE) of 25° between passive and active earplugs [127]. Custom molded earplugs increase sound localization performance over universal, foam earplugs by reducing the rate of front-back confusions from 24 % to 8 % [117].

Double protection reduces the number of correctly localized sounds from 70 % to 20 % [124], resp. from 95 % to 60 % [128]. Further, subjects which are using double protection testing several different head orientations before identifying the correct direction [118, 121].

Both, passive and active HPDs modify the user's individual cues and HRTF. Depending on the HPD type the modifications that are introduced in the user's natural HPD vary between 5 dB and 8.3 dB [127], between 1.4 dB and 4.6 dB [129], or between 1.3 dB and 7.6 dB [50].

Due to these concerns and complains about the comfort of HPDs the acceptance of HPDs decreases and the risk of hearing disorder increases [130, 108]. Builders, e.g. on constructions sites, need to protect them self from high level sounds and simultaneously need to notice the direction of warning sounds to avoid accidents [97]. Furthermore, dismounted soldiers are wearing HPDs for communication and protection purpose while they must perceive and analyze their acoustical environment fast and reliably [131]. Therefore, for certain users there is a conflict between the protection of their hearing and the ability of avoiding life-threatening situations by detecting and localizing them acoustically.

There are numerous studies about the influence of HPDs on the sound localization performance with respect to the HPD model [50, 117, 121, 119]. On the contrary, only few propositions have been published featuring HPDs which are preserving the directional information of the incoming sounds, allowing enhanced sound localization, and providing a natural perception of the surrounding sounds. The integration of a simplified outer ear geometry in the shell of earmuff hearing protection has been realized but no reliable evaluation concerning localization performance has been published [132]. As certain workers are exposed to conditions where hearing protection and sound localization are required simultaneously, and little research has been done yet the following work investigates how these two

requirements can be combined and how such HPDs may be designed.

1.4 Facilities

Investigations of HRTF and sound localization performance require appropriate equipment to obtain reliable data. In the following, the tools and setups used for data acquisition are presented.

1.4.1 Acquisition of HRTFs

Several databases of measured HRTFs have been published by researcher teams from different countries, e.g. the CIPIC database (USA) [61], the IRCAM database (France) [133], the ARI database (Austria) [134], and the ITA database (Germany) [135]. Each database is based on an individual format and representation style. The measurement conditions are slightly different between the groups. Some teams measured in the near-field others in the far field, some measured the Head-Related impulse response (HRIR) others the HRTF, and some measured the HRTF without others with hearing aids. The individual formats and measurement conditions limit the comparison of data from different databases. Spatially Oriented Format for Acoustics (SOFA)⁴ defines a file format for interchanging data related to spatial acoustics, e.g. HRTFs, Binaural room impulse responses (BRIRs), or Directional room impulse responses (DRIRs) [136]. SOFA has been accepted as standard AES69-2015 [137].

A board overview of measuring and acquiring HRTFs is presented in [138] with focusing on different individualization techniques. The advantage of such individualization techniques is that there is no need for HRTFs measurements. HRTF individualization tunes measured, non-individual HRTFs to a subject by comparing anthropometric data or by a perceptual selection process [138]. The underlying HRTF databases are often those mentioned above. These methods make it possible to identify an individualized HRTFs that is close to the subject's own HRTF. Nevertheless however they remain non-individual HRTFs. For the purpose of this work, different sets of individual HRTFs must be measured and simulated.

⁴Initiation of the project in Jan. 2012 and release SOFA 0.3 (1st documented release) in May 2013 [136].

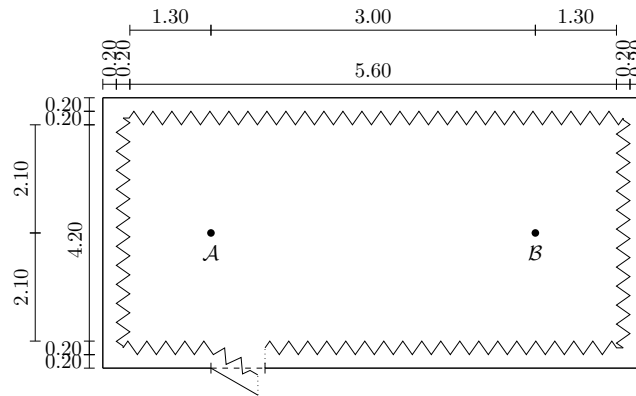


Figure 1.14: Floor plan of the anechoic chamber (inner room of the room within a room construction) with dimensions in meter. The chamber’s height is 1.80 m. Zigzag lines illustrate the acoustic absorbers. The measuring positions \mathcal{A} and \mathcal{B} are marked with dots.

1.4.1.1 Acoustic measurement

1.4.1.1.1 Anechoic chamber HRTFs are free field functions [38], so all HRTF measurements in the scope of this work were performed in a full-anechoic chamber. The chamber is a room within a room construction, entirely equipped with acoustic foam absorbers with triangular profile surface of 0.40 m height. With the tips of the absorbers defining the room’s boundary, the room’s dimensions are 4.20 m \times 5.60 m \times 1.80 m (W \times L \times H) and the room’s volume is 42.34 m³. The floor plan of the anechoic chamber is shown in Figure 1.14. Two positions, called \mathcal{A} and \mathcal{B} , are defined in the room. Both positions are at the room’s half height and the room’s half width. They are located at a distance of 1.30 m from the back or front wall of the room and at a distance of 3.00 m from each other.

Two series of each 32 cycles were performed to obtain the chamber’s reverberation time RT60. One cycle consisted of 1 s (1st series), resp. 2 s (2nd series) of pink noise followed by silence of the same length as the preceding noise. The noise signal was played back by a controlling computer, amplified by a MA1260 *Dayton* amplifier, and emitted through an *Electro-Voice* RX 115/75 loudspeaker at position \mathcal{A} . The reverberation time was measured during each cycle in position \mathcal{B} with an *NTi Audio* AL1 Acoustilyzer signal analyzer equipped with an *NTi Audio* MiniSPL measurement microphone. The signal analyzer measured accordingly to ISO 26101:2017 [139] at eight octave bands in the interval [62.5 Hz; 8 kHz] and averaged over all 32 cycles of one series. The averaged RT60 ranges between 0.24 s and 0.05 s, c.f. Table 1.3. The air temperature (21.4 °C and 21.5 °C), air humidity (31.3 % and 31.3 %), and atmospheric pressure (981.1 mbar and 981.0 mbar) were determined in the anechoic chamber using

1.4. FACILITIES

Frequency [Hz]	62.5	125	250	500	1000	2000	4000	8000
1 st series	0.25	0.20	0.16	0.07	0.06	0.05	0.05	0.06
2 nd series	0.24	0.20	0.17	0.06	0.06	0.05	0.05	0.05

Table 1.3: Reverberation time RT60 in seconds of the anechoic chamber, measured at eight octave bands between 62.5 Hz and 8 kHz.

a *Vaisala* WXT520 weather station before and after the RT60 measurement.

1.4.1.1.2 Setup The signal routing for the HRTF measurement setup was adapted from that of Gardner [140, 62]. A *Stanford Research* SR780 signal analyzer generated the test signal. It was amplified by a *Dayton* MA1260 power amplifier and emitted at position \mathcal{A} by a *JBL* Control 1 Pro loudspeaker, whose frequency response and impedance are shown in Figure 1.15. The acoustic signal was recorded in position \mathcal{B} , either with a reference microphone (*B&K* Type 4192 microphone with *B&K* Type 2669 preamplifier) or a dummy head. The recorded signal was returned to the SR780 signal analyzer, which calculated the transfer function between its the output signal and the returned signal. This transfer function is called $H_{dh}(\phi)$ with a dummy head involved and H_{mc} with the reference microphone involved. The signal routing for the measurement with the reference microphone is illustrated in Figure 1.16. The center of the loudspeaker was aligned with position \mathcal{A} and faced with its membrane towards position \mathcal{B} . Either the capsule of the reference microphone or the CoH of the dummy head was aligned with position \mathcal{B} . The dummy head was placed on a turntable allowing rotations in the horizontal plane. Figure 1.17 shows the measurement setup in the anechoic chamber with the dummy head on the turntable.

1.4.1.1.3 Dummy heads Dummy heads are preferred over human subjects, as the latter slightly move, tilt, or rotate their torso and heads. It has been proposed to either discard and repeat the measurement at the current direction of incidence if the subject has moved [142], or to accept the measurement and the containing deviations [143]. Mechanical fixation of the head prevents any movements [144, 145, 40]. Depending on the fixation apparatus, the measured HRTF may be influenced by the apparatus and the subjects' acceptance to participate in the measurement may be reduced. To avoid such problems and to ensure reproducibility, the measurements were performed on three dummy head configurations. In general, acoustic dummy heads are shaped like human heads and consist of ear simulators, ear canals and microphones. The microphones are positioned accordingly

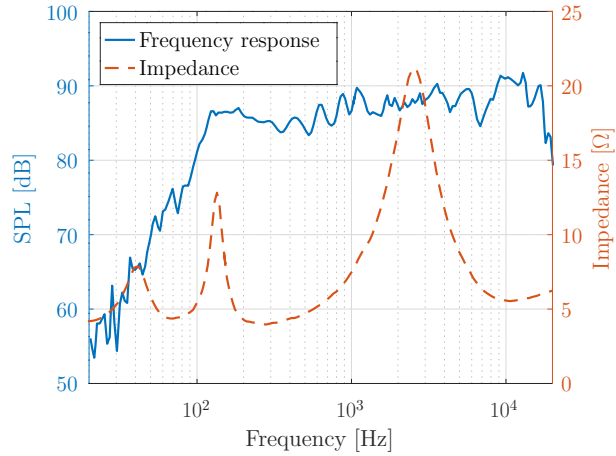


Figure 1.15: Frequency response and impedance of the loudspeaker model as specified in the manufacturer’s data sheet [141].

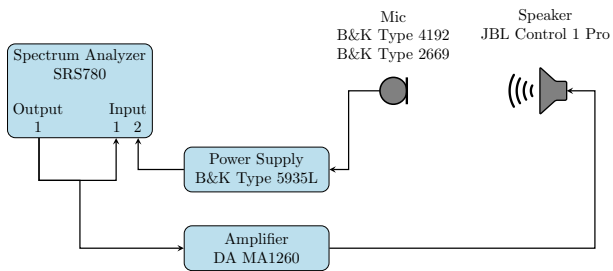


Figure 1.16: Configuration of the setup for measuring the transfer function of the measurement chain. For measuring HRTFs, the reference microphone was replaced by a dummy head.



Figure 1.17: View of the anechoic chamber with the loudspeaker on the left side in position \mathcal{A} and the dummy head on the turntable on the right side in position \mathcal{B} .



Figure 1.18: *ISL* dummy head equipped with Type 3.4 ear simulators, called Harry34.



Figure 1.19: *ISL* dummy head equipped with Type 3.3 ear simulators, called Harry33.



Figure 1.20: *B&K* head and torso simulator equipped with Type 3.3 ear simulators, called HATS33.

to the position of the human eardrum. The *French-German Research Institute of Saint-Louis (ISL)* dummy head, which has been originally designed to assess impulsive, high sound pressure levels [146], was used in two different configurations. Both configurations included *B&K* Type 4192 microphones and *B&K* Type 2670 preamplifiers. In the first configuration, called Harry34, the dummy head was equipped with *Head Acoustics* Type 3.4 ear simulators, c.f. Figure 1.18. In the second configuration, called Harry33, the dummy head was equipped with *Head Acoustics* Type 3.3 ear simulators, c.f. Figure 1.19. *B&K* Head and Torso Simulator Type 5128, c.f. Figure 1.20, was used as third dummy head configuration, in the following called HATS33. The ear simulators of Harry33 and HATS33 are designed accordingly to the recommendations given in [147]. Before returning the output signals of the dummy heads to the SR780 signal analyzer, they were processed by a *B&K* Type 5935L signal conditioner when using Harry34 or Harry33 and by a *B&K* Type 1704 signal conditioner when using HATS33.

1.4.1.1.4 Post processing The transfer function $H_{dh}(f, \phi)$ was measured at 16 angles of incidence in the horizontal plane from 0° to 337.5° in steps of 22.5° . Since the angle of incidence ϕ is defined from the view of the artificial head and this, not the loudspeaker, was rotated, the turntable had to

rotate to the angle $-\phi$. The test signal was a logarithmic increasing sine sweep ranging from 16 Hz to 25.6 kHz. Sweeps show high SNR, low distortion, and high robustness to time-varying side effects [148]. The sound pressure level at position \mathcal{B} was 94 dB SPL and 1600 frequencies ranging from 16 Hz to 25.6 kHz were measured.

The transfer function of the measurement chain $H_{mc}(f)$, c.f. Figure 1.16, was obtained similarly to $H_{dh}(f, \phi)$. Instead of using a dummy head, the reference microphone was mounted at position \mathcal{B} . $H_{mc}(f)$ is independent of the angle of incidence and is therefore measured only once.

The transfer functions $H_{dh}(f, \phi)$ were equalized with the inverse of $H_{mc}(f)$. The resulting HRTF

$$H_{HRTF}(f, \phi) = \frac{H_{dh}(f, \phi)}{H_{mc}(f)} \quad (1.5)$$

contains only head related information but still non-directional information such as the ear canal resonance [62]. The Directional Transfer Function (DTF) is obtained by zero averaging the HRTF over the measured angles of incidence [149]. The DTF contains only directional information as a function of the horizontal angle of incidence ϕ [150]. Only the DTF is considered in the following and from now on the term HRTF refers to the DTF.

1.4.1.2 Numerical simulation

The acoustic field around the object under test, e.g. a dummy head [151] or a human subject [152] can be calculated by numerical simulation. The simulation requires a 3D mesh of the object. Magnetic resonance imaging [153, 151, 66], stereo cameras [152, 154], structured light [155, 66, 156], or infrared light [154] technologies have been used to obtain the 3D mesh. To reduce complexity and fasten the calculation only the outer ear [154] or the head [157] has been considered, rather than the entire bust or body. With a low-cost *3D Systems* Sense 3D scanner, which is based on single band infrared light pattern technology [158], the dummy head configurations Harry34, Harry33, and HATS33, c.f. Figures 1.18 to 1.20, were scanned. The scanned meshes were post processed which included denoising, smoothing, and the verification that they are single component, closed and entirely manifold [159]. In order to obtain reliable results, numerical simulation requires a minimum edge length of $\frac{\lambda}{6}$, with λ denoting the wavelength [65]. Regarding the highest frequency of interest of 25.6 kHz, we used a constant edge length of 2.23 mm for all meshes.

In general, the acoustic field has been simulated by solving the three-dimensional Helmholtz equa-

tion around the 3D object by deploying the Boundary Element Method (BEM) [65] or Finite Element Method (FEM) [160]. Numerical simulations in the scope of this work were performed on the high-performance computing machine at Structural Mechanics and Coupled Systems Laboratory (LMSSC) of Conservatoire national des arts et métiers (Cnam) and on the ISL computing hub. Both were running implementations of solving the Helmholtz equation, once proposed by Gumerov [65], once by Ziegelwanger [159, 66]. Both implementations are BEMs coupled with the Multi-Level Fast Multipole Method (MLFMM) [161] and take advantage of the acoustical principle of reciprocity. Due to the principle of reciprocity the simulation time is independent to the number of angle of incidences [64]. The implementation of Ziegelwanger is an open-source project called Mesh2HRTF⁵.

1.4.2 Virtual acoustic environment

Sound localization performance can be assessed with VAEs. They can be generated either with an array of loudspeakers mounted in an (semi-) anechoic chamber or over stereo headphones [162].

The loudspeaker array has been arranged cubically [163, 50], spherically [121, 128], or circularly [119, 120]. A semi-circular array has been fixedly mounted in the horizontal plane [117], the median plane [164], or it can be pivoted to span a sphere [63]. The number of loudspeakers varied largely from 8 to 43 for a circular 2D setup and reached up to 277 for a spherical 3D setup [120, 117, 118]. For single loudspeaker setups, the loudspeaker has been moved dynamically with a robotic arm [165]. Real sound sources have been used exclusively [50, 120], or in combination with virtual sound sources by source panning [166].

For the headphone based VAEs, no special equipment is required apart from a rendering software, often running on a Personal Computer (PC), and the headphones [162].

Throughout literature, the stimuli sound has been a 0.2 s to 1 s lasting white noise, pink noise, sine tone, or tonal tone at a fixed level between 65 dB SPL and 75 dB SPL or at a listener dependent level between 20 dB SL⁶ and 35 dB SL [118, 119, 166, 167, 124, 168, 169].

The listeners had to turn their heads [165, 170] or steer a laser beam [171] in the perceived direction, shout out a loudspeaker number [117, 172] or log in their answers by pressing buttons on a handheld device [59, 50], or indicating the response on a touch screen [166] or a panel [120] placed in front of

⁵The project is hosted at <http://mesh2hrtf.sourceforge.net/>. Accessed on May 14, 2020.

⁶Sensation Level (SL): sound level above a listener's individual threshold of hearing [39].

them.

1.4.2.1 Initial setup

The initial setup of the VAE was originally presented in [166]. Eight circularly arranged loudspeakers with a spacing of 45° were mounted in a semi-anechoic chamber (Audiometric booth by *IAC Acoustics*). The walls and ceiling of the chamber are covered by acoustic knob absorbers with a total height of 0.05 m. The floor of the chamber is constructed of reflection-reducing material with flat surface. With the tips of the absorbers delimiting the chambers borders, its dimensions are $2.50\text{ m} \times 5.40\text{ m} \times 2.05\text{ m}$ ($W \times L \times H$) and the volume is 27.68 m^3 .

During sound localization tests the stimulus has been presented in total at 16 different directions. Eight directions were aligned each with the directions of the loudspeakers (sound sources that are placed there are real sound sources) and eight directions were aligned each with the center point of the pairs of two neighboring loudspeakers (sound sources that are placed there are virtual sound sources). The test was designed as forced-choice test. The listening subjects indicated the perceived direction on a touch screen device by pressing buttons, each corresponding to one of the sixteen directions.

1.4.2.2 Advanced setup

The aim of upgrading the VAE was to align with the 16 directions of the HRTF measurement without using virtual sound sources. This would have required many modifications of the initial setup in the semi-anechoic chamber. Therefore, a completely new setup was designed and installed to replace the initial setup. For the new setup, the number of loudspeakers of the initial setup was doubled and professional loudspeakers were installed. In its final stage the new setup supports multi source scenarios with independent source directions. In the following, details about the installed hardware and software setup, as well as important specifications are presented.

1.4.2.2.1 Reverberation Time The reverberation time RT_{60} of the semi-anechoic chamber was measured, following the protocol as already used for the RT_{60} measurement of the full anechoic chamber, c.f. Section 1.4.1.1. The measured reverberation time RT_{60} of the semi-anechoic chamber ranges between 0.29 s at 62.5 Hz and 0.06 s at 8 kHz, c.f. Table 1.4. Before, resp. after the reverberation time measurement, the air temperature (22.7°C , resp. 23.0°C), air humidity (997.6 mbar, resp. 997.4 mbar),

1.4. FACILITIES

Frequency [Hz]	62.5	125	250	500	1000	2000	4000	8000
1 st series	0.29	0.22	0.10	0.09	0.09	0.07	0.07	0.06
2 nd series	0.28	0.23	0.10	0.09	0.09	0.07	0.07	0.07

Table 1.4: Reverberation time RT60 in seconds of the semi-anechoic chamber for octaves between 62.5 Hz and 8 kHz.

and atmospheric pressure (45.8 %, resp. 44.5 %) in the semi-anechoic chamber were determined.

1.4.2.2.2 Loudspeaker transfer function The transfer functions of the 16 *JBL* Control 1 Pro loudspeakers were examined. The loudspeakers were placed in the full-anechoic chamber one by one at position \mathcal{A} , while the reference microphone was placed at position \mathcal{B} . The transfer function of the entire measurement chain was measured with a linear sine sweep ranging from 16 Hz to 25.6 kHz⁷. The measured transfer functions are shown in Figure 1.21a. The frequency dependent standard deviation over the 16 transfer functions is shown in Figure 1.21b. The Root Mean Square (RMS) of the standard deviation is 2.46 dB, resp. 1.23 dB in the frequency interval [16 Hz; 25.6 kHz], resp. [100 Hz; 18 kHz]⁸. The impedance of the loudspeaker reaches a peak at around 130 Hz with a bandwidth of 28 Hz, c.f. Figure 1.15. The quality factor equals to 4.6 which seems to be too steep for the signal analyzer to adjust adequately. This results in a local minimum of the measured transfer function and a local maximum of the standard deviation at around 130 Hz. The standard deviation does not exceed 1.5 dB in the frequency range [64 Hz; 15 kHz]. In this interval the RMS of the standard deviation decreases to 0.75 dB.

1.4.2.2.3 Hardware configuration All 16 loudspeakers were mounted in the semi-anechoic chamber in a circular array with a diameter of 2.20 m and at a height of 1.385 m above the floor, see Figure 1.22a. A black, acoustically transparent curtain was placed 0.20 m in front of the loudspeakers as shown in Figure 1.22b. Once the subject was seated at the listening position in the center of the array, he or she no longer saw the loudspeakers. Only a yellow cross, which is marked on the inside of the curtain in positive X direction, indicated the reference direction to the listener.

The loudspeakers were driven by 2 twelve channel *Dayton* MA2560 power amplifiers. Loudspeaker

⁷We wanted to cover the audible frequencies between 20 Hz and 20 kHz. Due to technical limitations of the signal analyzer only frequencies to the power of 2 or 10 were available. So for the lower, resp. upper frequency limit we used $16 = 2^4$, resp. $25600 = 10^2 \cdot 2^8$.

⁸Interval corresponds to the ± 3 dB frequency response range as given in the loudspeaker's data sheet [141]

1.4. FACILITIES

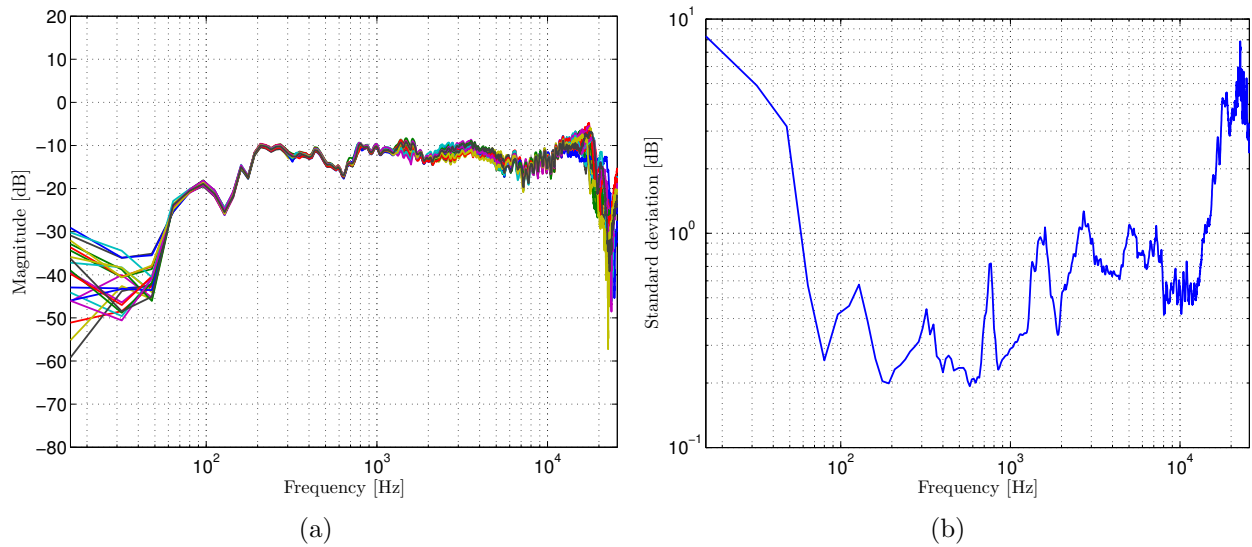
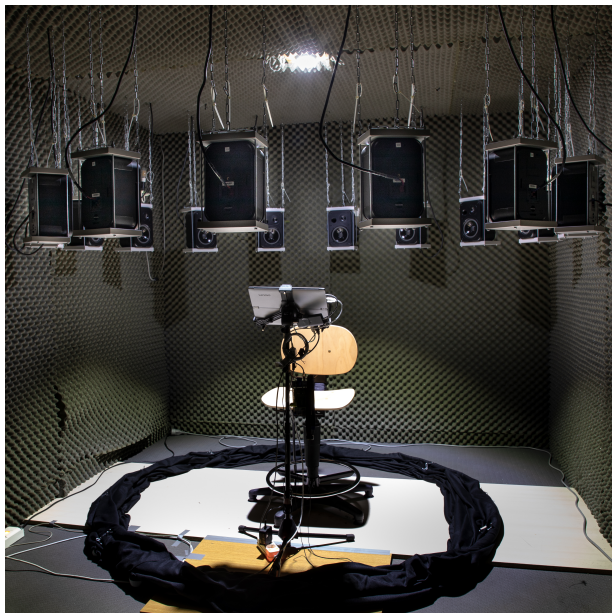


Figure 1.21: Influence of the loudspeakers on the transfer function of the measurement chain. Figure 1.21a shows the 16 transfer functions and Figure 1.21b shows the standard deviation of these 16 transfer functions on double logarithmic axis.



(a) Curtain was placed on floor for shooting.



(b) Montage image with curtain in place.

Figure 1.22: Setup of the listening test in the semi anechoic chamber including the circular loudspeaker array, the acoustically transparent curtain, the listening position, and the tablet computer in front of the listening position.

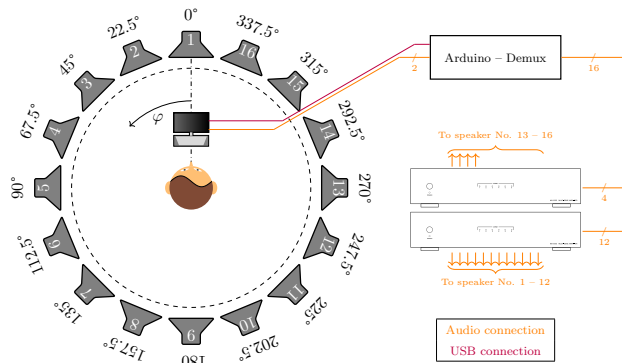


Figure 1.23: Sketch of the sound localization setup and the signal routing. Parallel, analog signals are marked with a slash on the signal line. The number of parallel signals is indicated by the digit next to the slash. The curtain is illustrated by the dashed circle between the circular array of sixteen loudspeakers and the listening position at the array’s center.

No. 1 to No. 12, resp. No. 13 to No. 16 were connected to channel 1 to 12, resp. 1 to 4 of the first, resp. second amplifier. The inputs of the amplifiers were fed by the outputs of 2 eight channel demultiplexers. The input of the demultiplexers were connected to the stereo audio output of the controlling computer. The first demultiplexer received the computer’s left audio signal, sending it to loudspeakers of odd numbers, and the second demultiplexer received the computer’s right audio signal, sending it to loudspeakers of even numbers. The demultiplexers were controlled by an *Arduino* MEGA2560 microcontroller platform, receiving the identifiers of the loudspeakers to be activated. These identifiers were sent via a digital serial connection from the controlling computer to the microcontroller platform. This signal routing is sketched in Figure 1.23.

1.4.2.2.4 Controlling software The controlling software comprised the functionalities for generating the VAE, placing the sound source at the angle of incidence ϕ , and the functionalities for conducting sound localization tests and providing user interaction. This software ran on a *Lenovo* Miix Idealpad 510, which served at the same time as the user interface. The sound of the sound source was defined in a mono audio file.

Sound source positioning The 2D vector base amplitude panning from [173] was adapted to place the sound source in the horizontal plane. The horizontal domain is split into 16 equal circular sectors $\mathcal{S}_i, i = 1, \dots, 16$, each delimited by two neighboring loudspeakers. Loudspeakers of sector \mathcal{S}_i are called \mathcal{L}_i^A , resp. \mathcal{L}_i^B and positioned at the angles $\varphi(\mathcal{L}_i^A)$, resp. $\varphi(\mathcal{L}_i^B)$ with $\varphi(\mathcal{L}_i^A) < \varphi(\mathcal{L}_i^B)$. The sector \mathcal{S}_i

in which ϕ falls is determined such that

$$\varphi(\mathcal{L}_i^A) \leq \phi < \varphi(\mathcal{L}_i^B) \quad (1.6)$$

is respected. Each sector has its own system of coordinate, which is aligned so that its X-axis points exactly between the two loudspeakers. The loudspeakers are located symmetrically around the X-axis at the angles 11.25° and -11.25° . The angle of incidence ϕ is transformed from the global system of coordinate to the angle of incidence $\phi_{\mathcal{S}}$ in the system of coordinate of the current sector \mathcal{S} by following Equation (1.7).

$$\phi_{\mathcal{S}} = (\phi \bmod 22.5) - 11.25 \quad (1.7)$$

The remaining step of the panning algorithm equals that presented by Pulkki in [173]. The gain vector $\mathbf{g} \in \mathbb{R}^{2 \times 1}$ for the two speakers is calculated by solving

$$\mathbf{p} = \mathbf{L}\mathbf{g} \quad (1.8)$$

where $\mathbf{p} \in \mathbb{R}^{2 \times 1}$ defines the position of the sound source and $\mathbf{L} \in \mathbb{R}^{2 \times 2}$ the position of the loudspeakers, both in Cartesian coordinates of the sector's system of coordinates. The audio signal from the mono audio file is duplicated and \mathbf{g} is applied to this stereo signal. The amplitude panned stereo signal is sent to the computer's stereo audio output. The identifiers of the delimiting loudspeakers \mathcal{L}_i^A and \mathcal{L}_i^B are sent to the microcontroller platform. The microcontroller platform routes the stereo audio signal to the dedicated pair of loudspeakers in the semi-anechoic chamber. This software design is independent of the sound in the mono audio file. The choice of sound, e.g. noise or speech, is related to the design of the individual listening tests and not limited by this setup.

Modes The Graphical User Interface (GUI) of the controlling software depends on the mode of operation of the controlling software. These modes of operation were defined accordingly to different phases of listening tests, and are called “adaptation phase”, “training phase”, and “testing phase”. The subject interacts with the software by using the touchscreen of the controlling computer. The common elements of the operation modes are a progress bar, a yellow cross (indicating frontal direction like that on the curtain), a black circle (indicating the acoustic horizon in the horizontal plane) and a black rayon (used as pointer). During the adaption phase the subject can freely position a sound source around him or her by steering the pointer in any direction, c.f. Figure 1.24a. The sound source

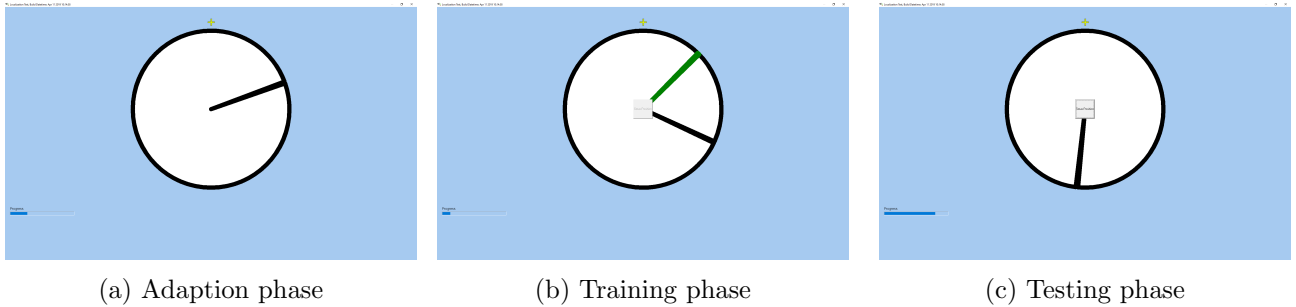


Figure 1.24: GUI of the controlling software in three different modes of operations, providing user interaction in combination with the touchscreen of the controlling computer.

in the VAE follows accordingly. During the training phase an additional button in the center of the black circle is available and sound is presented at random angles of incidence to the subject. Subjects steer the pointer in the direction where they perceived the recently presented sound and validate their choice by pressing the center button. The actual direction of the recently presented sound is indicated by an additional green rayon, c.f. Figure 1.24b. Before presenting the next sound, the green rayon is hided. The testing phase equals the training phase, except that the feedback with the green rayon is no longer provided, c.f. Figure 1.24c.

1.4.2.2.5 Sound source independence A limitation of the realized upgrade was the constraint that no more than one pair of loudspeakers could be simultaneously active. Hence in a multi-source scenario all sources would have been positioned in the same circular sector. This could be solved by a second upgrade. The microcontroller platform was replaced by a *MOTU 24Ao* audio interface, providing 24 analog audio output channels. The interface was connected to the computer via USB, making the stereo audio connection obsolete. The output channels 1 to 12 of the audio interface were connected with input channels 1 to 12 of the first amplifier and output channels 13 to 16 of the audio interface were connected with input channels 1 to 4 of the second amplifier. The MOTU AVB ASIO driver manages the communication on hardware level and allows full control over all output channel independently. The driver provides buffers, each of which is assigned to an output channel and can be accessed sample by sample. The controlling software was accordingly adapted to this new hardware configuration. The major change is that the identifiers of \mathcal{L}_i^A and \mathcal{L}_i^B are used to write the audio data into the buffers of the driver corresponding to the delimiting loudspeakers.

1.4. FACILITIES

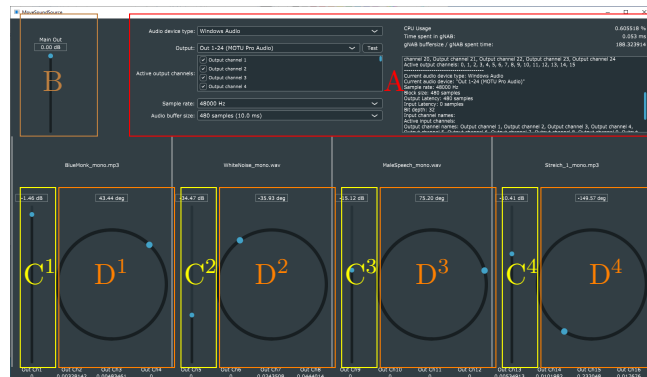


Figure 1.25: User interface of the demonstration software with 4 sound sources. Configuration of the hardware (A), control of the main output gain (B), and control of the gains (C^1 to C^4), the angles of incidence (D^1 to D^4) of source 1 to 4.

1.4.2.2.6 Demonstration software In addition to the controlling software a demonstration software was realized. This software takes the advantage of the individual control on each loudspeaker and works with the same hardware setup as the second upgrade of the controlling software. The demonstration software allows to position individually four sound sources at any angle of incidence in the horizontal plane. It is possible to increase the number of sound sources, which is limited only by the available computational power. The user interface of the demonstration software is shown in Figure 1.25, comprising controls for the gains (main and source gains) and the positions of the sound sources.

Chapter 2

Interaction of hearing protection and sound localization

Content

2.1	Hearing conditions	66
2.2	Head-related transfer function	67
2.2.1	Measurement setup	67
2.2.2	Obtained HRTFs	67
2.2.3	Changes in the HRTF	68
2.3	Localization test	78
2.3.1	Setup	78
2.3.2	Evaluation metadata	79
2.3.3	Localization results	82
2.4	Conclusion	94

It is frequently reported that the use of HPD interacts with sound localization, see Section 1.3.4.4. To develop the next generation HPDs that provide spectral cues for better sound localization, it is important to gain insight about HPD-related modifications in the HRTF and the influence of HPDs on sound localization performance. In this chapter we make use of the previously presented facilities. We measured the natural HRTF and the HPD modified HRTF of dummy heads in the anechoic chamber and conducted a subjective sound localization test with and without HPDs in a loudspeaker array based VAE.

2.1 Hearing conditions

Both, the HRTF measurements and the localization test were conducted under five hearing conditions. Each hearing condition is defined by the HPD under test. The natural hearing condition without any HPD is also referred to as protection null (P0). The non-natural hearing conditions are defined by the following HPDs. *ISL* “Bang”: an active, earplug HPD with universal 3-flange earplugs, referred to as protection one (P1), c.f. Figures 2.1a and 2.1b. Custom molded earplugs for P1 were available for dummy heads Harry34 and Harry33. P1 with custom molded earplugs is referred to as protection one custom (P1C). Figure 2.1c shows the custom molded earplugs for Harry33. *ZTac* “Z111”: an active, earmuff HPD, referred to as protection two (P2), c.f. Figure 2.2. *Nacre* “QuietPro”: an active, earplug HPD with universal foam earplugs, referred to as protection three (P3), c.f. Figure 2.3. *3M* “X5A”: a passive, earmuff HPD, referred to as protection four (P4), c.f. Figure 2.4.

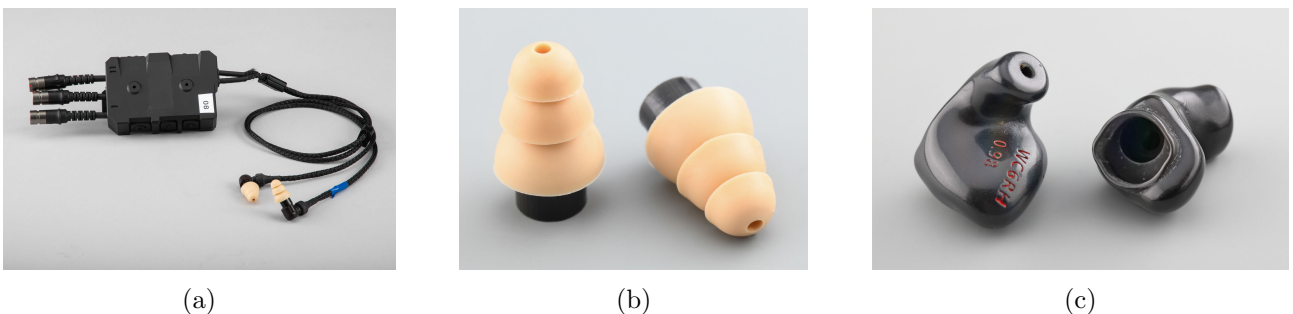


Figure 2.1: HPD P1 (Figure 2.1a) and detailed views of the flange earplugs (Figure 2.1b) and custom molded earplugs (Figure 2.1c). HPD P1, resp. HPD P1C refers to the HPD equipped with the flange earplugs, resp. custom molded earplugs.



Figure 2.2: HPD P2.



Figure 2.3: HPD P3.



Figure 2.4: HPD P4.

2.2 Head-related transfer function

2.2.1 Measurement setup

The HRTF measurements were performed in the anechoic chamber of *ISL*, as presented in Paragraph 1.4.1.1.1, . The HRTFs were measured in the horizontal plane at sixteen angles of incidence, constantly spread by 22.5° , ranging from 0° to 337.5° . The measurements were performed on three dummy heads, c.f. dummy head configurations presented in Paragraph 1.4.1.1.3, each under the five hearing conditions P0 to P4. Additionally, the HRTFs of Harry34 and Harry33 with hearing condition P1C were measured. The natural, i.e. open ear, HRTF of each dummy head was measured twice, while non-natural HRTFs were measured once. In total 20 HRTFs were obtained. The HRTF, resp. DTF, were calculated accordingly to the algorithm described in Paragraph 1.4.1.1.4.

2.2.2 Obtained HRTFs

The magnitude of the obtained HRTFs of the left ears are shown for each of the hearing conditions P0 to P4 for dummy head Harry34 in Figure 2.5, for dummy head Harry33 in Figure 2.6, and for dummy head HATS33 in Figure 2.7. The HRTFs of dummy heads Harry34 and Harry33 under the hearing condition P1C are shown in Figure 2.5. The angle of incidence ϕ (ordinate) is shown over the logarithmic scaled frequency (abscissa) in all images. The magnitude of the HRTF is color coded.

Noise-like characteristics below 50 Hz and narrow peaks across the entire horizontal plane around 130 Hz appear independently of the hearing condition and the dummy head. These effects are due to the transfer function of the measuring chain with a limited frequency band, c.f. Figure 1.21a, and the impedance peak of the loudspeaker at around 130 Hz, c.f. Figure 1.15.

Regarding the loudspeaker's ([100 Hz; 18 kHz]), resp. the measured ([16 Hz; 25.6 kHz]) frequency range, the RMSEs between the two natural HRTFs of the dummy head are 1.38 dB, resp. 2.16 dB (Harry34), 0.93 dB, resp. 2.00 dB (Harry33), and 0.37 dB, resp. 0.94 dB (HATS33). These RMSEs are inversely proportional to the repeatability of the measurements. The RMSE and poor repeatability for Harry34, and in some degree also for Harry33, are suspected to be caused by the top-heavy measurement setup. Rotating between two measurement positions causes extreme low frequency oscillation of the entire setup, which lead to positioning errors of the dummy head during the measurement. Reinforcing the stand of the setup and shortening the distance between the turntable and the dummy head will resolve this issue. HATS33 was mounted with less adapter connections and closer to the turntable, making the setup much more rigid and oscillation free.

The head shadow effect is clearly noticeable in the HRTF plots, c.f. Figures 2.5 to 2.7. Especially for the natural HRTFs the amplitude for ipsilateral positions ($0^\circ < \phi < 180^\circ$) is larger than the amplitude for contralateral positions ($180^\circ < \phi < 360^\circ$). The maxima and minima are reached at lateral positions. Magnitudes of 0 dB are obtained near median positions. The difference between the maximum and minimum magnitude at frequencies around 100 Hz converges towards 0 dB, while it reaches up to 50 dB for frequencies around 10 kHz. This value exceeds the interval $[-30 \text{ dB}; 30 \text{ dB}]$ determined for the ILD in Section 1.2.2.2. Different frequency scales lead to this difference. Additionally, the bright spot effect is best visible for frequencies around 1 kHz. There, the HRTFs show local maxima at $\phi = 270^\circ$ compared to the neighboring angles of incidence of $\phi = 247.5^\circ$ and $\phi = 292.5^\circ$.

2.2.3 Changes in the HRTF

Changes in the HRTF caused by the deployed HPD are noticeable by comparing Figures 2.5 to 2.8. Visual inspection carries out noise-like characteristics for frequencies above 16 kHz over the entire horizontal plane. We notice a strong attenuation of these high frequencies by all HPDs. This highly reduced the SNR during the measurements, causing the background noise of the measurement chain to be measured instead of the test signal. The vibro-mechanical characteristics of the HPDs lead to high attenuation in the high frequencies and only little attenuation in the low frequencies, i.e. the HPDs dampen high frequencies well and low frequencies little. For active HPDs, the transfer functions mainly of the microphones and loudspeakers together with the sampling rate define the bandwidth of the electro-acoustic system, which can limit the transmitted frequency range [96, 174, 175, 176].

2.2. HEAD-RELATED TRANSFER FUNCTION

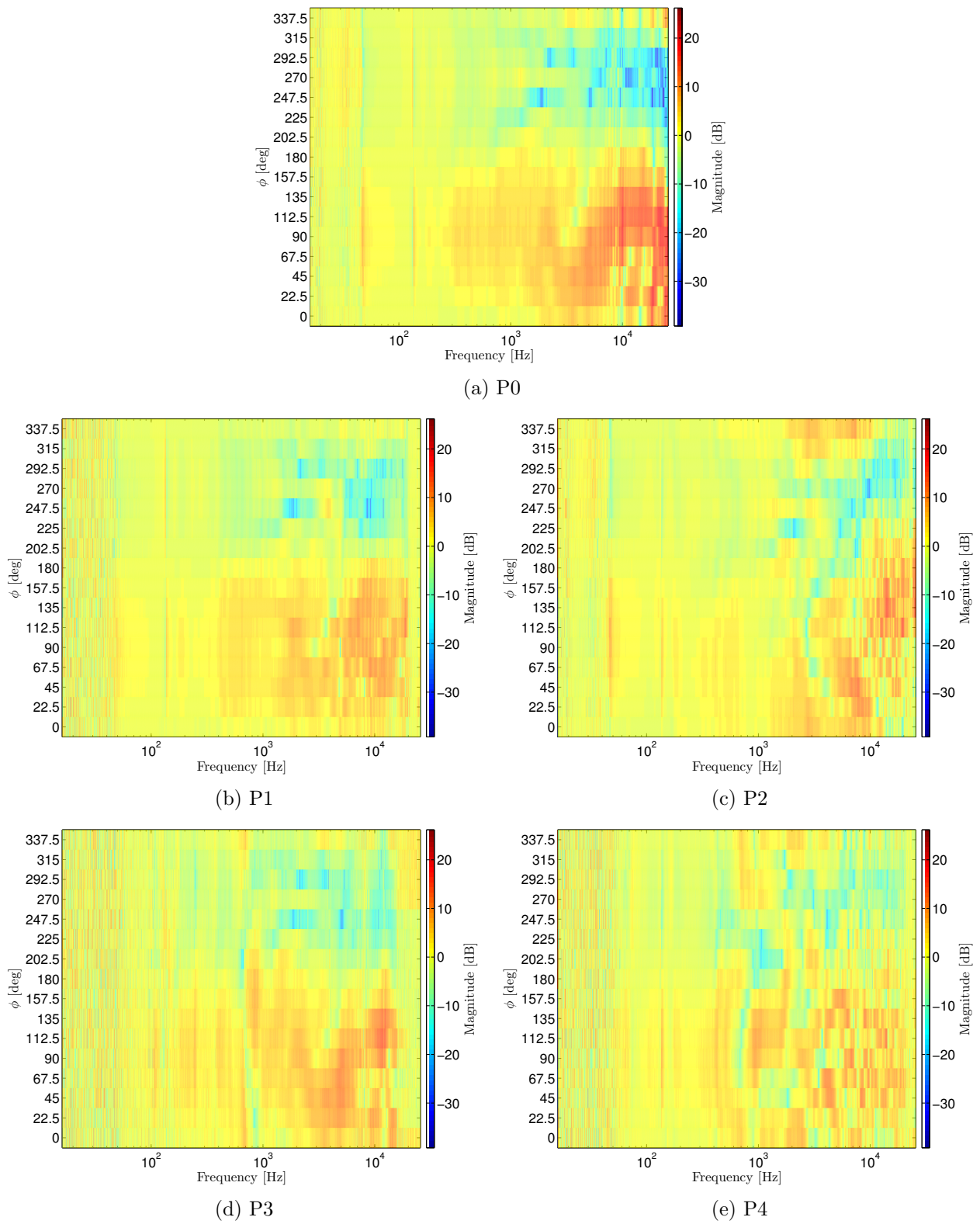


Figure 2.5: HRTFs of Harry34 with HPD P0 to P4.

2.2. HEAD-RELATED TRANSFER FUNCTION

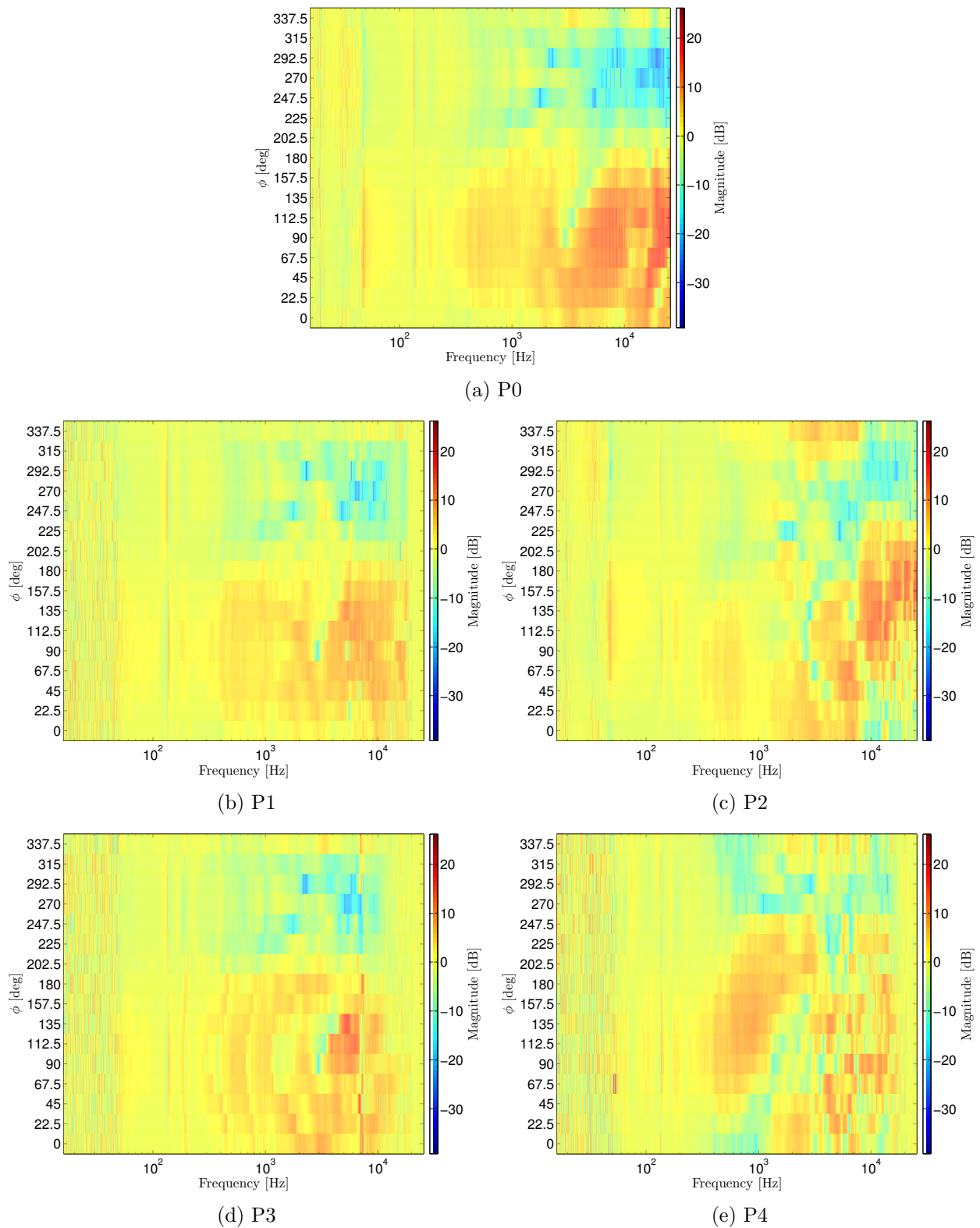


Figure 2.6: HRTFs of Harry33 with HPD P0 to P4.

2.2. HEAD-RELATED TRANSFER FUNCTION

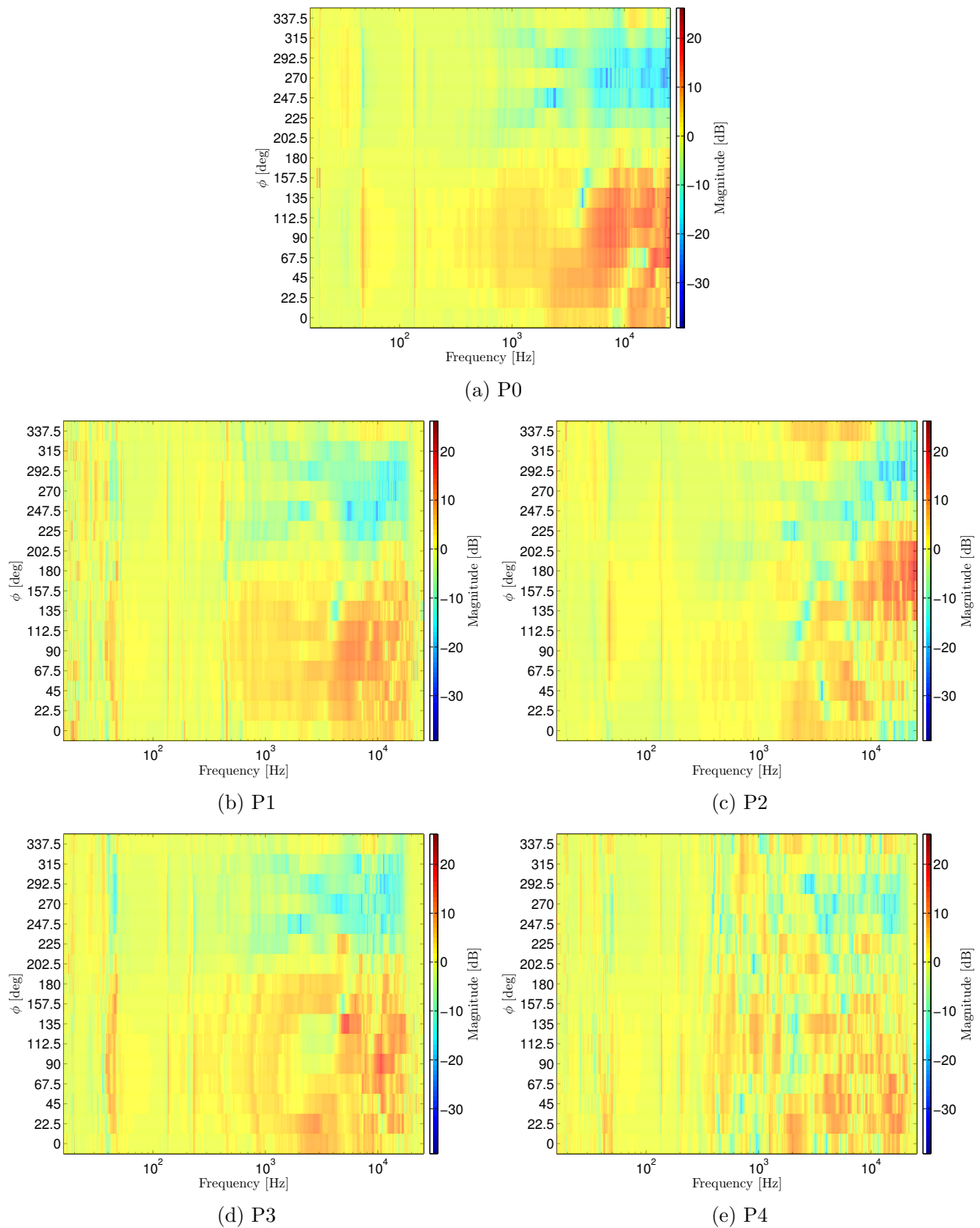


Figure 2.7: HRTFs of HATS33 with HPD P0 to P4.

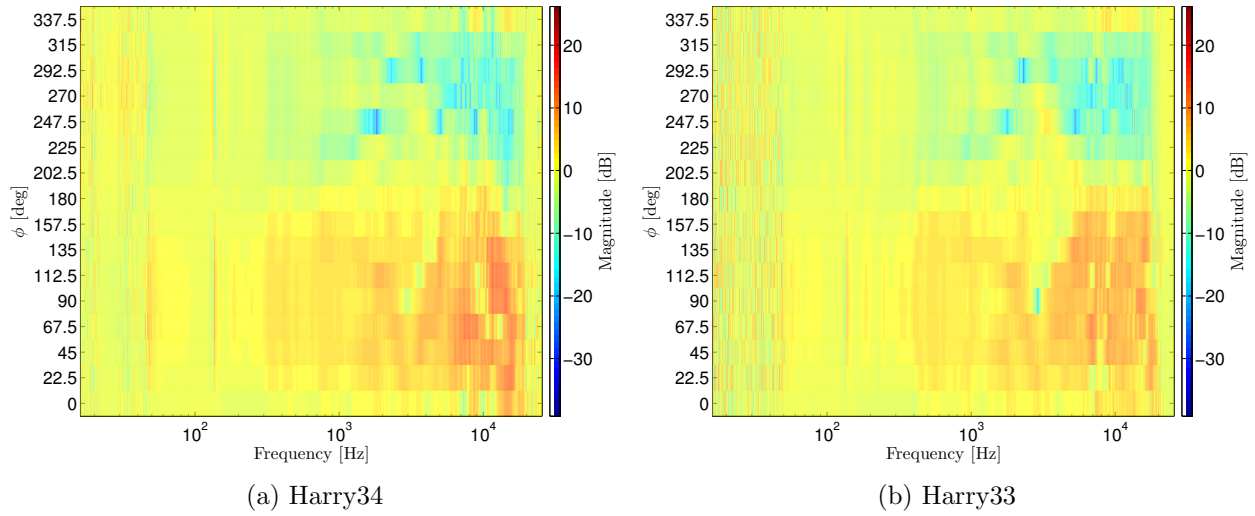


Figure 2.8: HRTFs of Harry34 and Harry33 with HPD P1C.

HPD P1 and P1C conserve the location, i.e. the position on the frequency axis, of notches and peaks of the original HRTF. In contrast, the magnitude of the notches and peaks are diminished, c.f. Figures 2.5b, 2.6b, 2.7b and 2.8. HPD P2 leads to a triangular shaped notch between 2.6 kHz and 6 kHz, symmetrically positioned around 90° with its tip located near 112.5° and 2.6 kHz c.f. Figures 2.5c, 2.6c and 2.7c. This symmetrical notch is most distinguished on the ipsilateral side and results in equal spectral cues for sounds coming from the frontal and back. Even though HPD P3 introduces modifications which are greater than those due to HPD P1, it still keeps cues of the original HRTF, c.f. Figures 2.5d, 2.6d and 2.7d. Major modifications are caused by HPD P4 over the entire frequency range which are leading to a strong degradation of the HRTF, c.f. Figures 2.5e, 2.6e and 2.7e.

2.2.3.1 Frequency banks

The auditory system focuses on frequency bands to evaluate the incoming spectrum, c.f. Sections 1.2.1 and 1.2.2.4, hence the changes in the HRTF were also analyzed in frequency bands. Seven filter banks were defined and used in the following: The filter banks named “all” and “speaker” each consisted of one frequency band, which were equal to the frequency ranges of the sine sweep (“all”) and the loudspeaker (“speaker”). The filter bank named “critBands” contained 24 frequency bands which correspond to Zwicker’s critical bands [12]. The frequency banks named “dir90”, “dir95”, “boost90”, resp. “boost95” contained 4, 2, 5, resp. 5, frequency bands which corresponded to Blauert’s directional and boosted bands of 90% and 95% confidence [40]. Figure 2.9 provides visual comparison of the

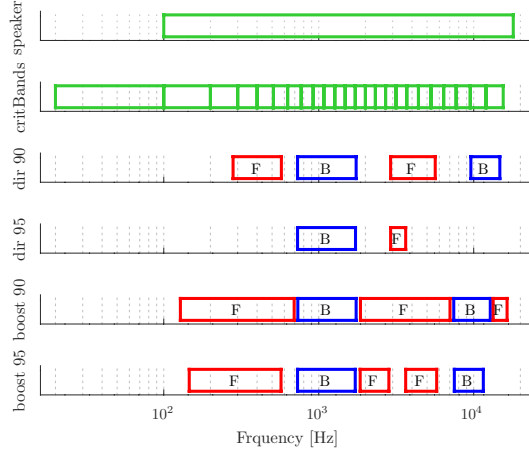


Figure 2.9: Filter banks used for the calculation of the RMSE. Filter bank “all” corresponds to the abscissa (frequency axis). Frequency bands, which induce a preference in the perceived direction, are marked with letter F (pure tones in these bands are primarily perceived in the frontal hemisphere) and letter B (pure tones in these bands are primarily perceived in the back hemisphere).

filter banks, their individual frequency bands and the covered frequency ranges.

2.2.3.2 HPD induced modifications

The modifications in the magnitude of the HRTFs were assessed on the pairwise RMSE between the natural hearing condition ($\text{HRTF}_0(f, \phi)$) and each of the non-natural hearing conditions ($\text{HRTF}_{P_i}(f, \phi)$, $i = 1, \dots, 4$), c.f. Equations (2.1) and (2.2). The RMSE $D(f)$ was calculated for each frequency band of a filter bank, followed by averaging over the frequency bands of the filter bank, resulting in the final RMSE of that filter bank. By this we obtained a rating which was based on a single numerical value. The RMSE values of the individual frequency bands of the “critBands” filter bank are kept for a more detailed investigation.

$$E(f, \phi) = 20 \cdot \log_{10} \left(\left| \frac{\text{HRTF}_{P_i}(f, \phi)}{\text{HRTF}_{P_0}(f, \phi)} \right| \right) \quad (2.1)$$

$$D(f) = \sqrt{\frac{1}{16} \cdot \sum_{\phi=0^\circ}^{337.5^\circ} E(f, \phi)^2} \quad (2.2)$$

Figure 2.10 visualizes the RMSEs between the P0 and the P1, P2, P3, and P4 hearing conditions based on the critical bands and for each dummy head respectively. Independently of the HPD and

2.2. HEAD-RELATED TRANSFER FUNCTION

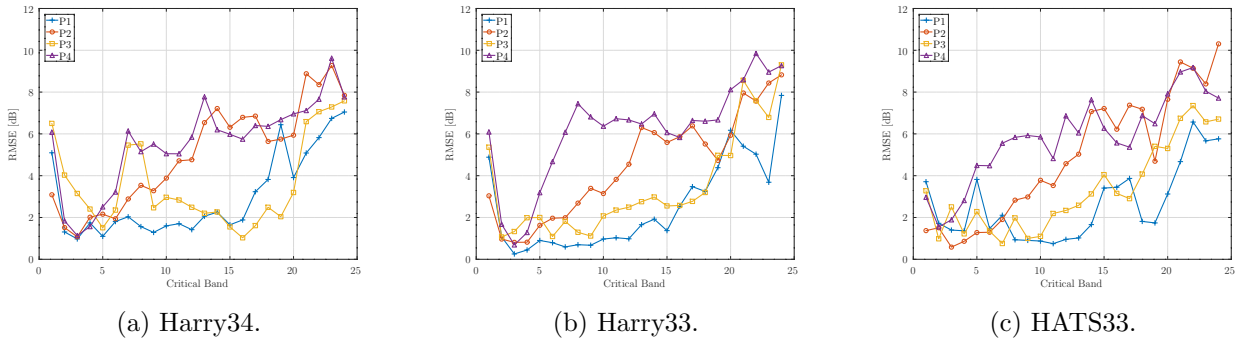


Figure 2.10: Root Mean Square Error (RMSE) between pairs of P0 and P1, P2, P3, resp. P4 in 24 critical bands after Zwicker [12], as obtained on three dummy heads.

the dummy head, high RMSEs were obtained in the low and high frequency range. The high values in critical band N° 1, i.e. for frequencies below 100 Hz, were caused by the limited bandwidth of the loudspeaker, c.f. Paragraph 1.4.1.1.1, leading to high variant results. The high RMSEs in the upper critical bands, i.e. high frequencies, were caused by the low pass characteristics of the hearing protections, which strongly dampen the high frequency components of the HRTF. The cutoff frequency and the edge steepness of the low pass characteristic depend on the hearing protection. The general trend of the RMSE in Figure 2.10 follows a steep decay from the 1st to the 2nd critical band followed by an increase which is characteristic for each HPD. The steep, resp. flat increase in the low frequencies is followed by a flat, resp. step increase in the high frequencies, see HPD P4 on Harry33 and HATS33, resp. HPD P3 on Harry34 and P1 on Harry33. Also a moderate increase over the entire frequency range can be noticed, see HPD P2 on HATS33. The RMSE does not exceed 4 dB until critical band N° 19 (Harry34), N° 19 (Harry33), and N° 21 (HATS33) for HPD P1. With the remaining HPDs, 4 dB are not exceeded until critical band N° 11, N° 12, and N° 12 (HPD P2), critical band N° 7, N° 19, and N° 15 (HPD P3), and critical band N° 7, N° 6, and N° 5 (HPD P4), respectively with dummy head Harry34, Harry33, and HATS33. Excluding the peak at critical band N° 7 and N° 8 of HPD P3 on Harry34, the RMSE does not exceed 4 dB until critical band N° 21. In general HPD P1 and P3 led to lower RMSEs between critical band N° 9 to 18 than HPD P2 and P4. The difference between the curves ranges from 0.2 dB in critical band N° 7 between HPD P3 and HPD P2 for Harry33 to 4.3 dB in critical band N° 13 between HPD P3 and HPD P2 for Harry34.

Figure 2.11 visualizes the RMSE, based on the single numerical values, between the natural hearing condition and the four non-natural hearing conditions for the three dummy heads. The RMSE was least

2.2. HEAD-RELATED TRANSFER FUNCTION

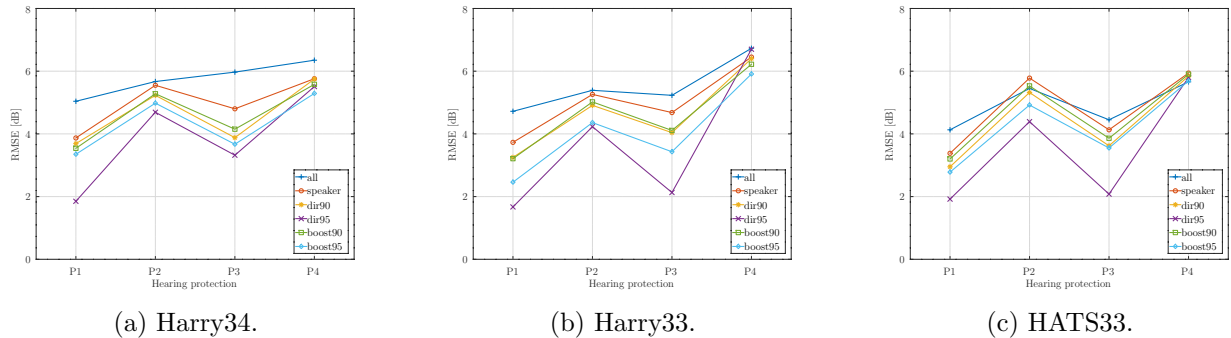


Figure 2.11: RMSE between pairs of P0 and P1, P2, P3, resp. P4 in six frequency bands as obtained on three dummy heads.

distinct across the hearing conditions in the frequency range of the sine sweep (c.f. filter bank “all”). The noisy signal outside the frequency range of the loudspeaker distorted the RMSEs. Focusing on the frequency range from 100 Hz to 18 kHz cuts out the non-reliable components of the HRTF outside this frequency range and hence led more distinct RMSEs between the hearing conditions. Three of Blauert’s filter banks exclude the 130 Hz and therefore the peak in the loudspeaker’s impedance. The fourth filter bank, i.e. “boost90”, has its lower limit at 130 Hz. Hence the RMSEs takes into account the impedance peak. Nevertheless, we found that the differences in the RMSE due to the hearing conditions were more important than whether a filter bank included the frequency of 130 Hz (“speaker” and “boost90”) or excluded (“dir90”, “dir95” and “boost95”) the frequency of 130 Hz.

It is obtained that filter banks of reduced total bandwidth result in reduced RMSEs, with exception of applying frequency bank “dir95” on the data of HPD P4. Further, they enlarge the RMSE differences between earplugs and earmuffs. With dummy head Harry34 the “speaker” filter bank results in RMSE of 3.87 dB, resp. 4.80 dB (P1, resp. P3) and 5.55 dB, resp. 5.76 dB (P2, resp. P4), while with the “dir95” filter bank the RMSEs are 1.85 dB, resp. 3.32 dB (P1, resp. P3), and 4.69 dB, resp. 5.51 dB (P2, resp. P4). The inter group distance between earplug HPDs (group G1) and earmuff HPDs (group G2) is calculated with the adaption of the single-linkage function from [177], c.f. Equation (2.3). $RMSE(x, fbk)$ returns the RMSE value of the protection x for the filter bank fbk . The return values of $RMSE(x, fbk)$ correspond to the visualized data in Figure 2.11.

$$D_{G1,G2}(fbk) = \min (|RMSE(x, fbk) - RMSE(y, fbk)|) \forall x \in G1, y \in G2 \quad (2.3)$$

The inter group distance between earplug and earmuff HPDs increases from 0.75 dB (filter bank

2.2. HEAD-RELATED TRANSFER FUNCTION

	P2	P3	P4
P1	✓,✓,✓	✗,✗,✗	✓,✓,✓
P2	–	✗,✗,✓	✗,✓,✗
P3	–	–	✓,✓,✓

Table 2.1: Acceptance (✓) or rejection (✗) at the 5 % level of the hypothesis that the pairs of HPD have RMSE data of different distributions. For dummy heads Harry34, Harry33, HATS33 respectively.

“speaker”) to 1.37 dB (filter bank “dir95”) for dummy head Harry34, from 0.58 dB (filter bank “speaker”) to 2.1 dB (filter bank “dir95”) for dummy head Harry33, and from 1.65 dB (filter bank “speaker”) to 2.31 dB (filter bank “dir95”) dummy head HATS33.

The ANalysis Of Variance (ANOVA) is applied on the RMSE data shown in Figure 2.11 with the groups P1, P2, P3, and P4. The p-values are $3.26 \cdot 10^{-4}$ (Harry34), $6.64 \cdot 10^{-6}$ (Harry33), and $2.61 \cdot 10^{-7}$ (HATS33). Hence, for all dummy heads, the hypothesis is rejected at the 5 % level that the RMSE data are based on distributions with the same mean. Consequently, there is at least one group which is statistically significant different than the others. The result of a pairwise ANOVA test at the 5 % level, testing the hypothesis that the RMSE data are of different distributions, is summarized in Table 2.1. It turns out that for all dummy heads the earplug HPDs (P1 & P3) introduce significantly less modifications to the HRTF than the passive earmuff HPD, as well as HPD P1 introduces less modifications than P2, while P1 and P3 do not lead to significant different modifications. For two out of three dummy heads the modifications of P2 are not significantly different than those of P3 nor P4. This dependence on the dummy head seems to be caused by the fitting of the HPD on the dummy head and the coverage of the outer ears. HPD P2 has small ear cups that make perfect covering of the outer ears very difficult, while HPD P4 has large ear cups that do not cause this difficulty. The varying fit of HPD P2 and the slightly different covering of the outer ears cause these head-related dependencies.

In the beginning of this section, it was visually carried out that the earplug HPDs P1 and P3 introduce less modifications in the HRTFs than the earmuff HPDs P2 and P4. The filter bank analysis based on the RMSEs between the natural and non-natural hearing conditions proves this visually identified characteristics. Ordering the considered HPDs by increasing RMSE, i.e. by increasing modifications on the HRTF, gives the series P1, P3, P2, and P4. Additionally, earplug HPDs introduce significantly less modifications on the HRTF than earmuff HPDs.

2.2. HEAD-RELATED TRANSFER FUNCTION

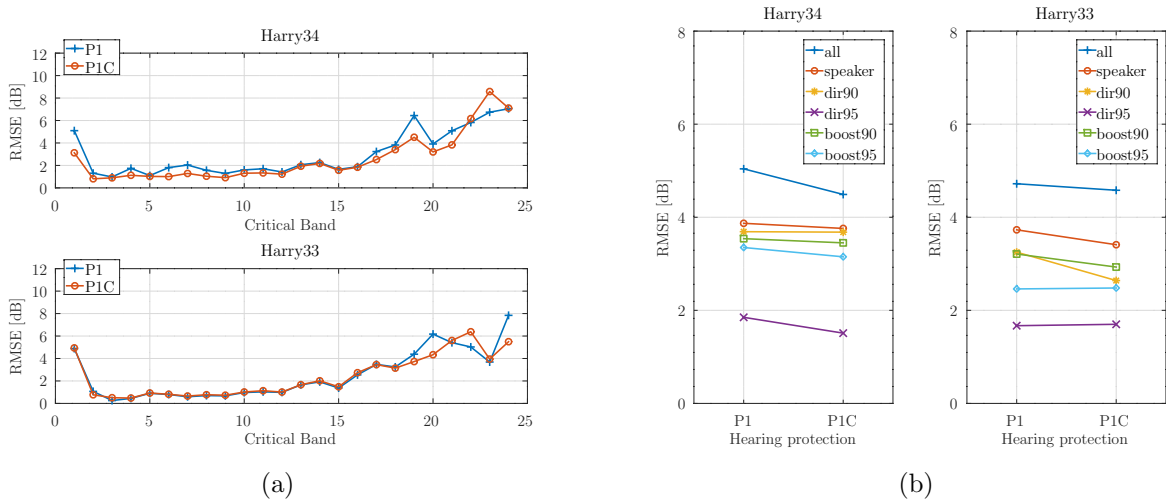


Figure 2.12: RMSE between pairs of P0 and P1, resp. P1C in 24 critical bands after Zwicker [12] (Figure 2.12a) and in six frequency bands (Figure 2.12b), as obtained on dummy head Harry34, resp. dummy head Harry33.

2.2.3.3 Earplug induced modifications

Figure 2.12 compares the effects on the HRTF when using a universal (3-flange) earplug (c.f. HPD P1) and a custom molded earplug (c.f. HPD P1C) on the dummy heads Harry34 and Harry33.

The RMSEs between HPD P0 and HPD P1 and P1C, respectively are shown in Figure 2.12a for the filter bank “critBands”. The curves for HPD P1 and HPD P1C show similar trends and are very close to each other (RMSE between the two curves: 0.83 dB (Harry34); 0.69 dB (Harry33)). Only in the high frequencies they show individual details. Using the custom molded earplug instead of the universal earplug reduces the RMSE in average by 0.40 dB (Harry34), resp. 0.10 dB (Harry33) over all critical bands. The custom molded earplug reduces the RMSE in critical band N° 1 to N° 21 (in average by 0.57 dB) on Harry34 and in critical band N° 2, N° 17 to N° 20, and N° 24 (in average by 0.88 dB) on Harry33. In critical band N° 22 to N° 24 (Harry34) and N° 1, N° 3 to N° 16, and N° 21 to N° 23 (Harry33), the custom molded earplug increases the RMSE in average by 0.75 dB (Harry34) and 0.17 dB (Harry33).

The RMSEs based on single numerical value are illustrated in Figure 2.12b for hearing condition P1 and P1C. The previously discussed characteristic, i.e. a reduced RMSE with reduced total bandwidth of the filter bank, reappear for both dummy heads. The customized earplugs reduce the RMSE in average by 0.22 dB for both dummy heads. The differences in the RMSE between the universal and

2.3. LOCALIZATION TEST

custom earplug vary between 0.55 dB (filter bank “all”) and 0.01 dB (filter bank “dir90”), resp. 0.61 dB (filter bank “dir90”) and -0.03 dB (filter bank “dir95”) for dummy head Harry34, resp. Harry33. Note that on Harry33 HPD P1C lead to a slightly increased RMSE for filter banks “boost95” and “dir95”.

An ANOVA on the RMSE data shown in Figure 2.12b with the groups P1 and P1C, result in p-values of 0.719 (Harry34) and 0.718 (Harry33). This rejects the null-hypothesis that the RMSE of P1 and P1C are significantly different, for both dummy heads. Hence, it cannot be confirmed that custom molded earplugs introduce less modifications in the HRTF than universal 3-flange earplugs. They might provide enhanced fitting and improved subjective comfort but should not be preferred nor spurned over custom molded earplugs concerning preservation of spectral cues.

2.3 Localization test

The previously measured modifications of the HRTF caused by HPD provide objective evaluation on the spectral cues but not on how subjects deal with these modified cues during sound localization. To assess this issue, the performance of localizing sound sources by human listeners wearing HPDs is examined in the following and compared to the performance obtained with natural hearing. For this purpose, a subjective sound localization test was conducted.

2.3.1 Setup

In addition to the actual sound localization, a Bekesy audiometry is part of this localization test. Both, the localization test and the audiometry, were conducted in the semi-anechoic chamber which is presented in Section 1.4.2.2.

The Bekesy audiometry was performed prior to the localization test, in order to determine the subjects’ hearing threshold at the left and right ear [178]. The hearing threshold data allow to identify if subjects have normal hearing or not, and if the sound localization performance is influenced by the individual hearing thresholds. A pulsed pure tone was used as stimuli. Seven frequencies between 125 Hz and 8 kHz, spaced by octaves were tested. The audiometry was conducted automatically using an *Otometrics* Madsen Astera² clinical audiometer with TDH39 headphones connected and controlled by the *Otometrics* Otosuite audiometry software [179].

During the localization test subjects were asked to indicate the perceived direction of sounds which

2.3. LOCALIZATION TEST

were presented at discrete directions of incidence $\phi_t = 0^\circ, 22.5^\circ, \dots, 337.5^\circ$. 16 different directions of incidence were considered. The test was designed as a non-forced choice test, consisting of five series. In each series, subjects performed the test under a different hearing condition. The subjects always performed the first series without HPD. The non-natural hearing conditions followed in series 2 to 5 with a random order of HPD P1 to P4, c.f. Section 2.1. Even though the order of HPD P1 to P4 was random, it was assured that when considering the set of all participants, each series consisted of an equal distribution of HPD P1 to P4.

Before the first series, subjects completed the adaption phase to familiarize with the haptic of the user interface and the test environment. During 30s they were placing the sound source freely around them. The entire set of subjects was split into two groups: the test group and the control group. Both groups consisted of an equal number of subjects. All subjects performed in all series the testing phase. Subjects of the test group performed an additional training phase prior to each testing phase. All 16 directions were trained twice and tested five times. The presentation order of the directions is random. Subjects in the test, resp. control group listen to 112, resp. 80 sounds. White noise of 200ms was chosen as test sound. The controlling software recorded the response time for each sound. The response time equals the length of the responding interval. The responding interval starts by the end of the recently presented sound and ends as soon as the subject logs in the response. The subjects were told to focus the yellow cross in front of them and not to move their head during listening. No mechanical fixture was used to ensure this constrain; instead, subjects were asked to put on a head-motion tracker. This self made head tracker device is based on an *Adafruit* “BNO055” Absolute Orientation Sensor and a *ST* NUCLEOF767ZI board. It recorded the heading angle ϕ_h (same conventions as for ϕ , c.f. Section 1.1) of the subjects’ heads during the experience at a sampling frequency of 50 Hz.

2.3.2 Evaluation metadata

In total 40 subjects with normal hearing, related to their age, participated in the listening test. 11 were female and 29 were male, 4 left-handed and 36 right-handed. The average age of the subjects was 31.45, with a standard deviation of $\sigma = 10.47$. The age of the subjects varied between 20 years and 57 years. We notice a large skewness towards younger ages as more than two third of the participants were younger than age 35. Figure 2.13 shows the age distribution of the subjects in the control group

2.3. LOCALIZATION TEST

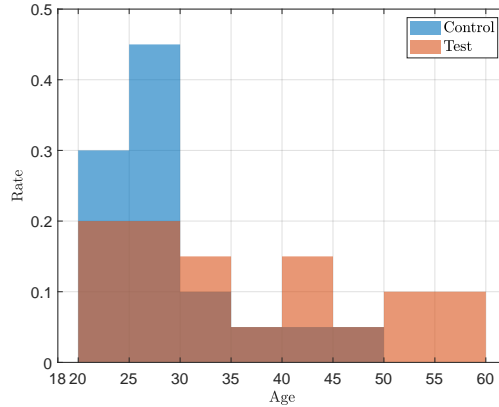


Figure 2.13: Distribution of the age of subjects of the control and test group.

and test group. The population of the control group is in average younger (mean: 29.25 years) than the population of the test group (mean: 36.10 years). Further, the age of control subjects spans a smaller interval (min age: 20 years, max age: 40 years) and shows a larger skewness towards younger subjects ($> 80\%$ are younger than age 35) than test subjects (min age: 20, max age: 57, $\approx 55\%$ are younger than age 35).

2.3.2.1 Response Time

The average response time is 2.42 s. The response times per hearing condition and group are summarized in Figure 2.14. For none of the hearing conditions P0 to P4 is there any difference in response time between trained and untrained subjects (p-values: $\gg 0.05$). The average response times under the natural hearing condition (HPD P0) are 2.79 s (test group) and 2.90 s (control group). Under non-natural hearing conditions the response times decrease at least by 0.51 s (test group) and 0.52 s (control group). This decrease is slightly significant in the test group for HPD P1 to P3 (p-values: < 0.05) but not for P4 (p-value: > 0.05). Regarding the control group, this difference is significant for earplug HPDs P1 and P3 (p-values: < 0.01) and little significant for earmuff HPDs P2 and P4 (p-values: < 0.05). The average response time varies under the non-natural hearing conditions between 2.25 s (HPD P3) and 2.35 s (HPD P4) in the test group and between 2.30 s (HPD P3) and 2.38 s (HPD P2) in the control group. These variations are not significant in the test group (p-values: $\gg 0.05$) nor in the control group (p-value: $\gg 0.05$).

The response time t_r averaged over all subjects and its moving average over 5 samples is confronted

2.3. LOCALIZATION TEST

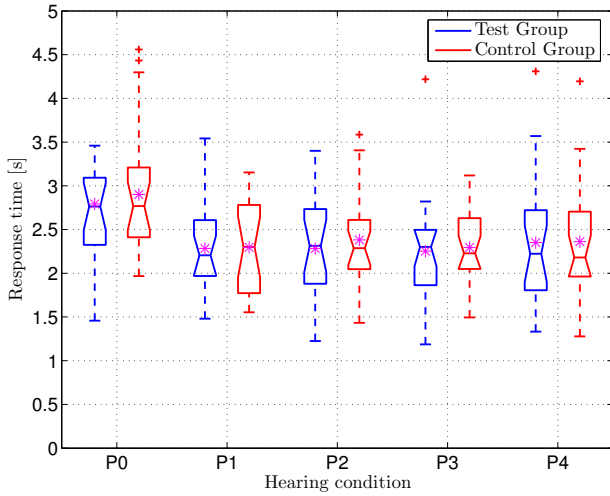


Figure 2.14: Box plot with mean values (asterisks) of the response time for training and control group HPD. Data are shifted horizontally for better visibility.

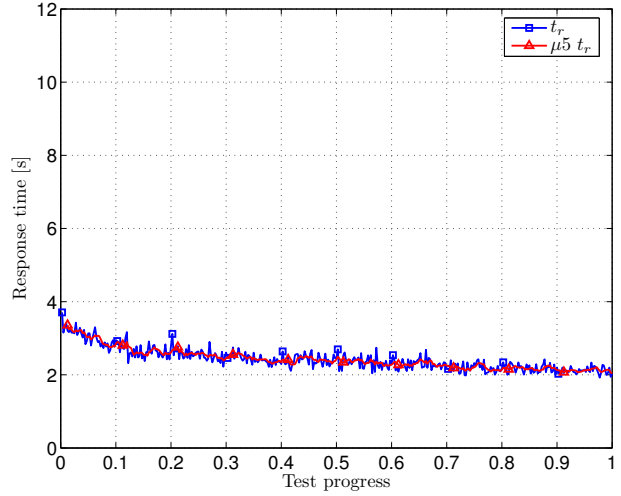


Figure 2.15: Response time t_r and its moving average over 5 samples $\mu_5 t_r$ as a function of the test progress.

with the test progress in Figure 2.15. At a test progress of 0 the test is about to start, at a test progress of 0.1 10% of the test progress are finished, and so on until a test progress of 1 where 100% of the test is finished, i.e. all 5 series are completed. Starting at 3.71 s (3.36 s for the moving average) both curves show the trend of the decreasing response time along test progress. It can be noticed that once 15% of the test is completed, the slopes of the curves decrease.

Remembering that all subjects started with HPD P0, subjects showed high motivation in order to perform well in the beginning of the test. As the progressed, subjects became familiar with both the test protocol and the user interface, making them accelerate. Moreover, the loss of motivation caused by the monotonous task and the length of the series lead them to finish the test faster. Since non-natural hearing conditions only occur at increased test progress, we obtained an inevitably reduced response time for the non-natural hearing conditions. Additionally, the deployed earplug HPDs are testing the correct fit after power-on. Subjects had to reinsert the earplugs several times before they were correctly in place, which annoyed and unsettled the subjects. This results in the slightly reduced response time with earplugs over earmuffs, most noticeable in the control group.

2.3.2.2 Head position

The head movements of the subjects are assessed by the RMS head position during the responding intervals. With a responds interval length of 2.42 s (c.f. average response time) and a head tracker sampling rate of 50 Hz, the RMS head position is calculated over 121 head position samples. Figure 2.16 visualizes the distribution of the RMS head positions ϕ_h during the responding intervals over the length of the responding intervals (equal to the response times). Stating the hypothesis that head rotations during the responding interval lead an increased RMS head position and increased response time, the distributions should show peaks clearly off the X- and Y-axis. As the obtained data is primarily distributed either along the X-axis or Y-axis it is concluded that subjects did not move their heads during the responding interval and hence respected the initial instructions. There are peaks in the distribution at $\phi_h \in [50^\circ; 60^\circ]$ and response times close to 2 s. These peaks are not due to head movements but to uncertainties of the head tracker on the head position. The head tracker determines the head position based on the geomagnetic field, which is, however, highly damped inside the semi-anechoic chamber due to its metal structure. Consequently, the head tracker losses the magnetic field and its reference orientation, which is leading to discontinuities in the heading data. For the individual hearing conditions, the RMS head position equals 14.61° (HPD P0), 7.95° (HPD P1), 13.37° (HPD P2), 8.75° (HPD P3), and 11.13° (HPD P4). We notice that the head movement is reduced for earplug conditions compared to the natural and earmuff conditions. An ANOVA shows that there is no statistically significant effect of the hearing condition on the head movement (p-value: $\gg 0.05$). The obtained varying head positions are related to the loss of the head tracker reference direction in the semi-anechoic chamber.

2.3.3 Localization results

The raw data of the localization stage are visualized with confusion diagrams in Figure 2.17. The three graphs represent the data, grouped accordingly to the type of hearing condition: natural hearing (HPD P0), hearing with earplugs (HPD P1 & P3), and hearing with earmuffs (HPD P2 & P4). The ideal localization performance, resp. ideal front-back confusion is illustrated by the main diagonal, resp. the anti-diagonals. The responded angles ϕ_r are discretized with 32 steps and a constant step width of 11.25° and are shown over the tested angle ϕ_t in steps of 22.5° . Each data point represents the probability with which ϕ_r falls into one of the 32 discretization intervals, as a

2.3. LOCALIZATION TEST

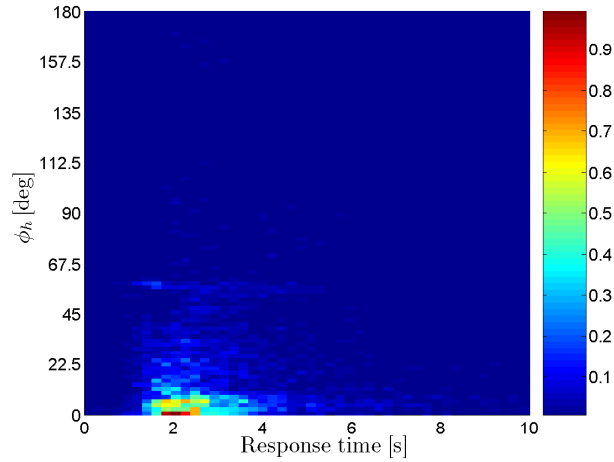


Figure 2.16: Distribution of the subjects' head positions ϕ_h and the response times. The obtained probabilities are color coded.

function of ϕ_t . Independently of the type of hearing condition, sounds on the median and frontal plane ($\phi_t \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$) are most often correctly perceived. Whereas the responded angle ϕ_r is most scattered across the entire horizontal plane for median positions ($\phi_t \in \{0^\circ, 180^\circ\}$). Lateral positions $\phi_t \in \{90^\circ, 270^\circ\} \pm 22.5^\circ$ are localized most frequent at 90° and 270° . This aligns with the general high localization performance for frontal sounds and poor performance for lateral sounds [4]. Further, most data points are in the lower left ($\phi_t \leq 180^\circ, \phi_r \leq 180^\circ$) or the upper right ($\phi_t > 180^\circ, \phi_r > 180^\circ$) submatrix of the confusion diagram. Hence, the most important localization confusions are front-back errors, nevertheless some sparse left-right errors and mixed errors occur. Despite these similarities, the confusion matrices show individual characteristics that depend on the hearing condition. The natural hearing, c.f. Figure 2.17a, shows perfect localization with a naturally occurring dispersion. Some few front-back confusions are obtained which generally occur at frontal directions, i.e. most of the front-back errors are caused by front-to-back errors. The absence of visual cues leads subjects to localize sounds with ambiguous cues behind them. The subjects' experience from their everyday life leads them to localize sound sources that are outside their visual field usually behind them. With earplugs (Figure 2.17b) and earmuffs (Figure 2.17b) sounds are less precisely located. Data are spread over the entire lower-left and upper-right quadrants of the confusion matrices, while earmuffs lead to slightly more left-right and mixed confusions than earplugs.

As we see from the confusion matrices, the localization performance varies with the hearing conditions. Therefore, the data of the localization test is examined in the following with respect to the

2.3. LOCALIZATION TEST

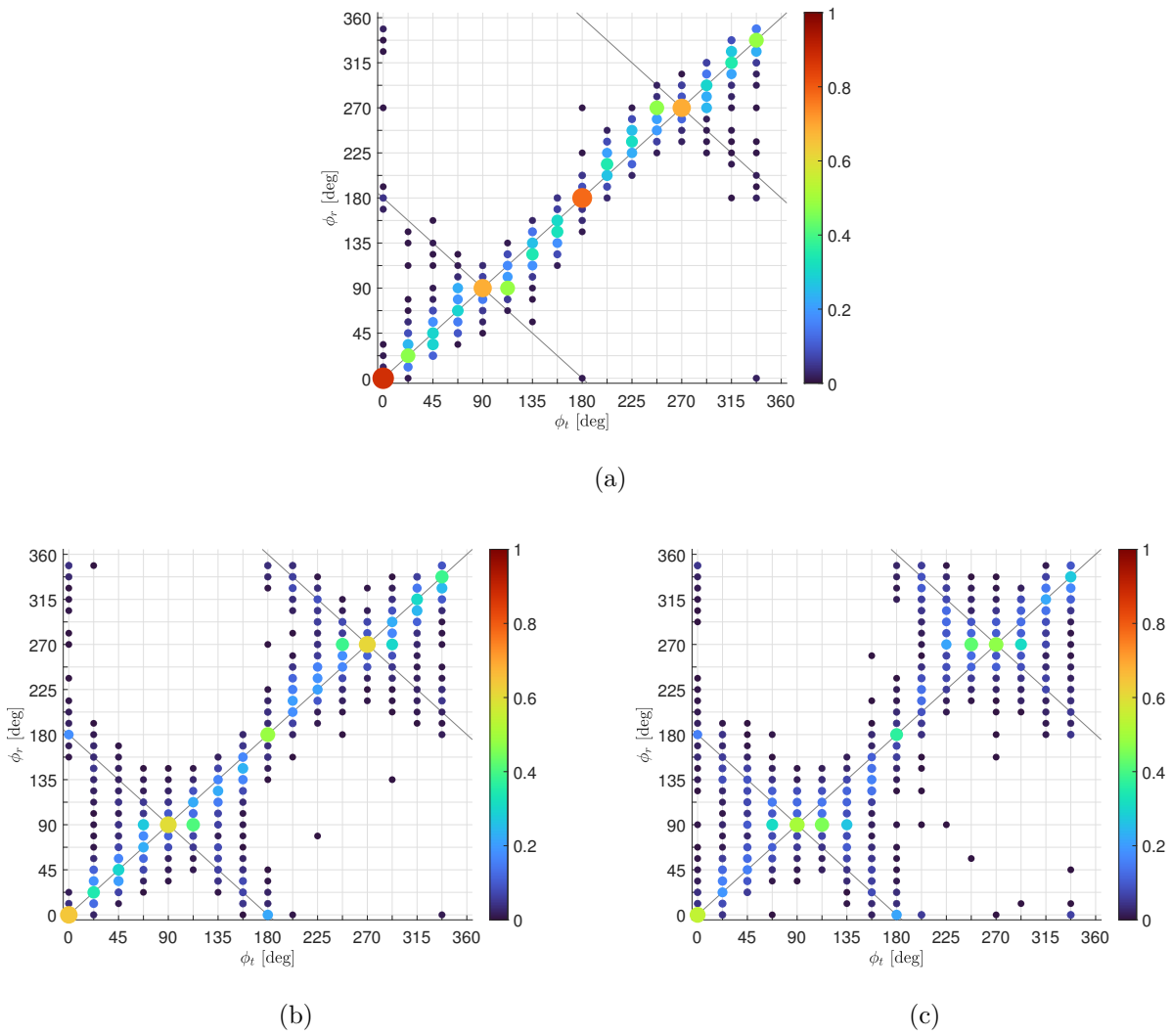


Figure 2.17: Confusion matrices under natural hearing (HPD P0, Figure 2.17a), under earplugs (HPD P1 and P3, Figure 2.17b), and under earmuffs (HPD P2 and P4, Figure 2.17c).

individual hearing conditions. Further, we will also evaluate whether sound localization is improved with training when HPDs are used.

2.3.3.1 Precision

The precision is in general a measure for the repeatability of a measurement. The conducted test is based on an unforced choice design, but stimuli are presented at a set of discrete directions. Therefore, we are interested in whether subjects recognize that sound is emitted from a limited number of distinct directions. We define the precision as a measure of how well the responded angle ϕ_r matches the discrete direction of the loudspeaker closest to ϕ_r , regardless of the actual direction ϕ_t of the sound and the error between ϕ_t and ϕ_r , c.f. Equation (2.4).

$$prec(\phi_r) = \left| \left(\frac{\phi_r}{\phi_s} \bmod 1 \right) - 0.5 \right| \cdot 2 \quad (2.4)$$

ϕ_s denotes the spacing of the loudspeakers of 22.5° . The precision is 1 for $\phi_r = n \cdot \phi_s$ with $n = \{0, 1, 2, \dots, 15\}$, that is, when ϕ_r points precisely in the discrete direction of one of the 16 loudspeakers. The precision is 0 for $\phi_r = (n + 0.5) \cdot \phi_s$ with $n = \{0, 1, 2, \dots, 15\}$, that is, when ϕ_r points perfectly in-between two loudspeakers.

The average precision is 59.47%. This value indicates that the responded angles have a mean offset of 6.69° from the discrete, angular positions of the loudspeakers. The precision varies between 59.89% (P2) and 61.80% (P0) in the test group and between 56.76% (P3) and 59.31% (P0) in the control group, c.f. Figure 2.18. The precision does not significantly differ between the hearing conditions in both groups (p-values: > 0.05). Training increases the precision in average by 2.13%, meaning that trained subjects localize sound sources 0.24° closer to the discrete, angular position of any loudspeaker than untrained subjects. Only for the natural hearing condition HPD P0 (p-value: < 0.05) and the non-natural hearing condition HPD P2 (p-value: 0.05), the effect of the training can be considered as significant. The precision increases due to training by 2.49% (P0) and 1.07% (P2), corresponding to an angular difference of 0.28° (P0) and 0.12° (P2) between trained and untrained subjects. These angular differences are below the minimum audible angle for frontal directions, varying between 1° and 3° [180] and below the angle which corresponds to the minimum audible ITD, c.f. Section 1.2.2.1. The differences obtained are not due to training, since they are below the possible performance of

2.3. LOCALIZATION TEST

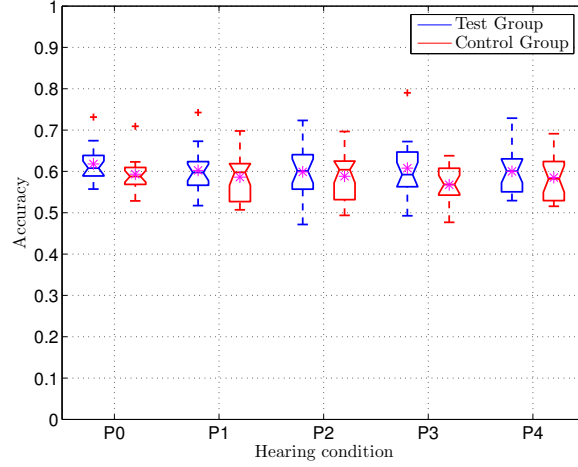


Figure 2.18: Box plot with mean values (asterisks) of the precision of the subjects responses on the loudspeaker positions.

the auditory system, but are due to statistical deviation. Although subjects in the test group get indications about the positions of the sound sources during training, it cannot be obtained that they remember these locations and localize more accurately during the testing phase than subjects of the control group without training. Additionally, it cannot be confirmed that trained subjects recognize the 16 different discrete angular directions of sound incidence, whereas untrained subjects do not.

2.3.3.2 Angular error

The signed, resp. unsigned angular error is defined in Equation (2.5), resp. Equation (2.6). Based on the symmetry along the interaural axis, ϕ'_r , resp. ϕ'_t is the projection of ϕ_r , resp. ϕ_t into the frontal hemisphere, c.f. Equation (2.7).

$$\epsilon(\phi_r) = \phi'_r - \phi'_t \quad (2.5) \quad |\epsilon(\phi_r)| = |\phi'_r - \phi'_t| \quad (2.6)$$

$$\phi'_\xi = \begin{cases} 180^\circ - \phi_\xi, & \text{if } 90^\circ < \phi_\xi \leq 180^\circ \\ 540^\circ - \phi_\xi, & \text{if } 180^\circ < \phi_\xi < 270^\circ \\ \phi_\xi, & \text{else} \end{cases}, \quad \xi \in \{r, t\} \quad (2.7)$$

The distribution of the signed angular error is visualized in Figure 2.19. The step size is 5.625° and equals to one quarter of the loudspeaker spacing. The maximum of the distribution is located at 0° with a height of 0.23. Hence, the signed angular error varies within $\pm 5.625^\circ$ for nearly a quarter of all data. In 59.91%, resp. 87.44% of all data the signed angular error does not exceed $\pm 11.25^\circ$, resp.

2.3. LOCALIZATION TEST

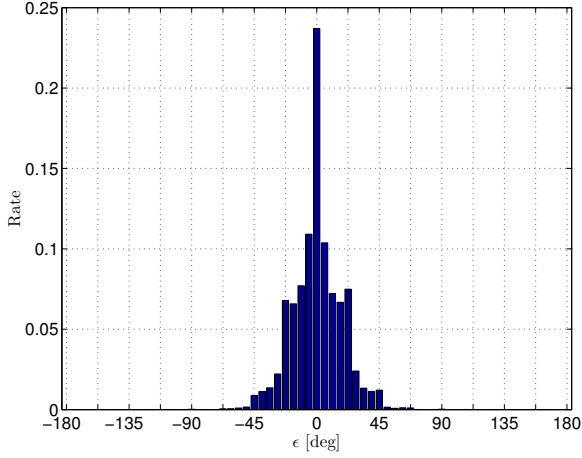


Figure 2.19: Distribution of the signed angular error ϵ with a resolution of 5.625° .

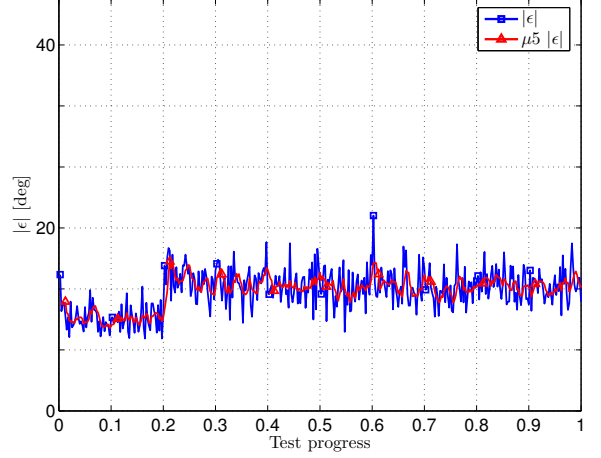


Figure 2.20: Unsigned angular error $|\epsilon|$ and its moving average over 5 samples $\mu_5 |\epsilon|$ over the test progress, both averaged over all subjects.

$\pm 22.5^\circ$. The distribution of ϵ decays towards 0 outside the $\pm 22.5^\circ$ range.

The trend of the unsigned angular error $|\epsilon|$ and its moving average over 5 samples is plotted over the test progress in Figure 2.20. During the first fifth of the test, i.e. 1st series with the natural hearing condition, the unsigned angular error is in average 11.16° . This agrees well with the overall error of 12° obtained with a 250 ms lasting noise in [181] and with the mean azimuth error of 13° obtained with a 750 ms lasting pink noise in [182]. During the last 80% of the test, the unsigned angular error is in average 13.69° . This value corresponds to the average value across all HPDs as the order of the non-natural HPDs P1 to P4 in the series 2 to 5 is random for each subject. The effect of the individual HPD P1 to P4 on the test progress is therefore not noticeable in Figure 2.20. The increased unsigned angular error during the last 80% of test shows that HPDs lead to a statistically significant (p-value: $\ll 0.001$) reduced localization performance with respect to natural hearing.

2.3.3.3 Correct responses

A response is defined as correct if the responded angle ϕ_r is within an interval of $\phi_i = 45^\circ$ (loose condition) or $\phi_i = 22.5^\circ$ (tight condition) around the tested angle ϕ_t , c.f. Equation (2.8).

$$cr(\phi_r) = \begin{cases} 1, & \text{if } \phi_r \in \left[\phi_t - \frac{\phi_i}{2}; \phi_t + \frac{\phi_i}{2} \right], \\ 0, & \text{else} \end{cases}, \quad \text{with } \phi_i = \begin{cases} 45^\circ & \hat{=} \text{ loose condition} \\ 22.5^\circ & \hat{=} \text{ tight condition} \end{cases} \quad (2.8)$$

2.3. LOCALIZATION TEST

Subjects responded in 65.51 % and 43.50 % with an angle considered correct under the loose and tight condition, respectively. The tight condition allows to compare our results with the results of forced-choice localization tests. It discretizes the responded angles from the continuous scale to the discrete locations of the loudspeaker.

Figure 2.21 shows the correct response for each HPD and group under the loose condition, c.f. Figure 2.21a, and tight condition, c.f. Figure 2.21b. Most correct responses are obtained with HPD P0: 87.75 %, resp. 84.64 % (loose condition) and 63.12 %, resp. 58.04 % (tight condition), in the test, resp. control group. Our results fall within the range of correct response rates that spans the range [59.88 %; 96.1 %] across the literature [126, 183, 119, 166]. The correct response rate decreases under non-natural hearing and ranges the interval [48.87 %; 76.31 %] (loose cond.), resp. [29.88 %; 54.25 %] (tight cond.). The correct response rate with earplugs is at least 18.06 %, resp. 11.25 % (loose cond.) or 16.63 %, resp. 8.21 % (tight cond.) higher than with earmuffs, respectively in the test and control group.

Focusing on the loose condition, pairwise ANOVA tests show that the correct response rate is significantly different for pairs of HPDs (p-values: $\ll 0.001$) in both groups. Except between the pairs of P1–P3, which are earplugs, and P2–P4, which are earmuffs, no significant difference is obtained (p-values: > 0.05). The difference between HPD P0 & P1 obtained in the test group is less significant (p-value: < 0.01) than with the control group (p-value: $\ll 0.001$). Same between HPD P3 & P4 where the difference between both HPDs is less significant in the control group (p-value: < 0.01) than in the test group (p-value: $\ll 0.001$).

Under the tight condition the difference between HPD P0 and P1 is less significant in the test group (p-value: < 0.05) than in the control group (p-value: $\ll 0.001$) while the difference between HPD P4 and P1, resp. HPD P4 and P3 is less significant in the control group (p-values: 0.0012, resp. $0.019 < 0.01$, resp. < 0.05) than in the test group (p-values: $\ll 0.001$).

The correct response rate increases due to training in average by 0.059 (loose cond.) and 0.057 (tight cond.). The effect of the training is significant under the loose, resp. tight condition for HPD P1 (p-value: < 0.05 , resp. < 0.01) and HPD P3 (p-values: < 0.01). For the natural hearing condition and earmuff hearing conditions there is no training effect obtained (p-values: > 0.05).

The trend of the correct response rate along the test progress, c.f. Figure 2.22, shows similar

2.3. LOCALIZATION TEST

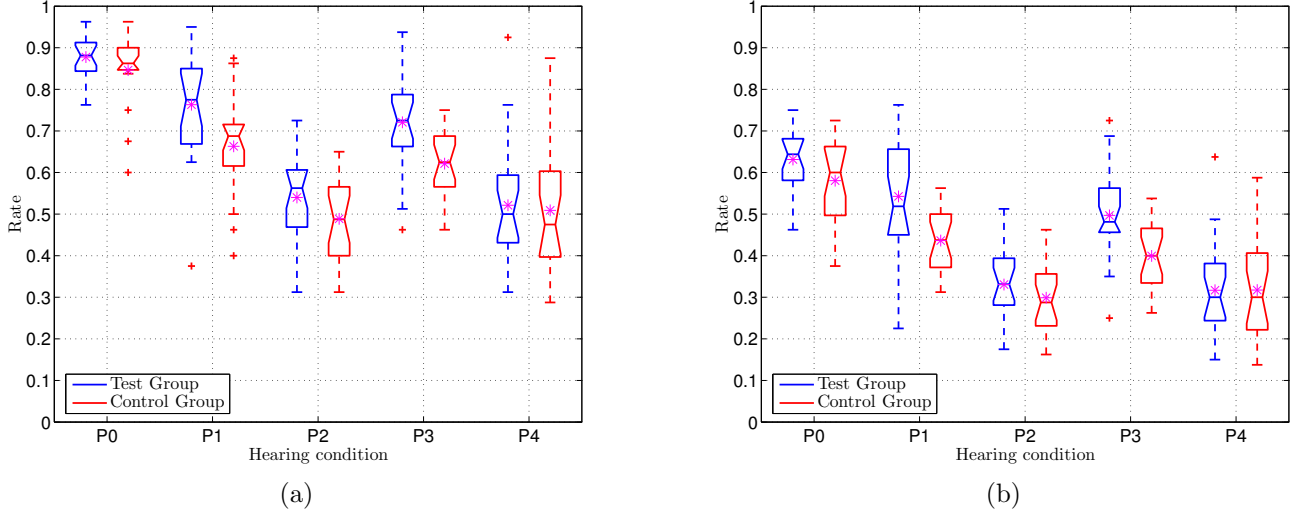


Figure 2.21: Box plots with mean values (asterisks) of the correct response rate for the loose condition ($\phi_i = 45^\circ$, Figure 2.21a) and the tight condition ($\phi_i = 22.5^\circ$, Figure 2.21b).

characteristics as already obtained for the unsigned angular error in Figure 2.20. Between 0% and 20% of the test progress the correct response rate is remarkably different than above 20% of the test progress. The correct response rate decreases when passing from the natural hearing condition to non-natural hearing conditions. The correct response rate is in average 81%, resp. 56% below 20% of the test progress and decreases to 60%, resp. 39% for the remaining test, for the loose, resp. tight condition.

2.3.3.4 Correct quadrant

The correct quadrant condition returns 1 if the responded angle ϕ_r and the corresponding tested angle ϕ_t are in the same quadrant, c.f. Equation (2.9).

$$cq(\phi_r) = \begin{cases} 1, & \text{if } \left\lfloor \frac{\phi_r}{90^\circ} \right\rfloor = \left\lfloor \frac{\phi_t}{90^\circ} \right\rfloor \\ 0, & \text{else} \end{cases} \quad (2.9)$$

The correct quadrant rate indicates the probability that data satisfy the correct quadrant condition. It is obtained by averaging the number of correct quadrants in a set of data. This set of data can represent the groups of trained and untrained subjects or the HPDs. The correct quadrant is not considering data where the tested angle ϕ_t is in either the median or frontal plan, c.f. Equation (2.10). These angles align with the bounds of the quadrants and are excluded to avoid miscounting of the

2.3. LOCALIZATION TEST

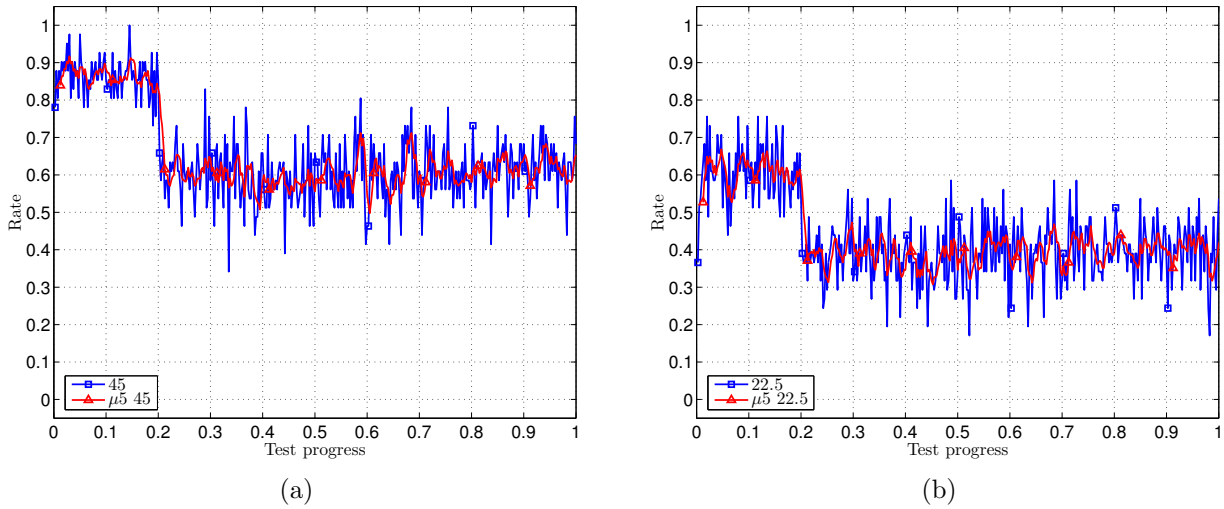


Figure 2.22: Trend of the correct response rate along the test progress for the loose condition $\phi_i = 45^\circ$ (Figure 2.22a) and the tight condition $\phi_i = 22.5^\circ$ (Figure 2.22b).

correct quadrant rate.

$$\phi_t = 90^\circ \cdot n, n \in \mathbb{N}_0 \quad (2.10)$$

Sounds are most often localized in the correct quadrant (test group: 93 %, control group: 90 %) with natural hearing, c.f. Figure 2.23. With earplug and earmuff HPDs, the localization performance based on the correct quadrant condition is equal to the localization performance obtained based on the two correct response conditions, c.f. Section 2.3.3.3. With HPD P1 (84.75 %, resp. 75.08 %) and P3 (78.50 %, resp. 68.97 %) higher performance is obtained than with HPD P2 (60.83 %, resp. 56.98 %) and P4 (59.42 %, resp. 58.49 %), respectively in the test and control group. The difference between natural and non-natural hearing is significant for all HPDs (p-values: < 0.001) in both groups. This significance between natural and non-natural hearing has also been found in [166]. This difference is less significant (p-value: < 0.01) in the test group between HPD P0 and HPD P1. The difference between earplug and earmuff HPDs is not significant (p-values: > 0.05) in neither the test nor the control group. The difference between HPD P2 and P3, resp. HPD P3 and P4 is less significant in the control group (p-values: < 0.01) than in the test group (p-values: < 0.001).

Training has no effect on the correct quadrant rate under the natural hearing (p-value: > 0.05), nor under the earmuff hearing (p-values: $\gg 0.05$). In the case of HPD P0, it is difficult to further increase the performance of 90 % correctly localized sounds. In contrast, HPD P2 and P4 highly degrade the spectral cues such that even performing a training stage with those HPDs does not lead to improved

2.3. LOCALIZATION TEST

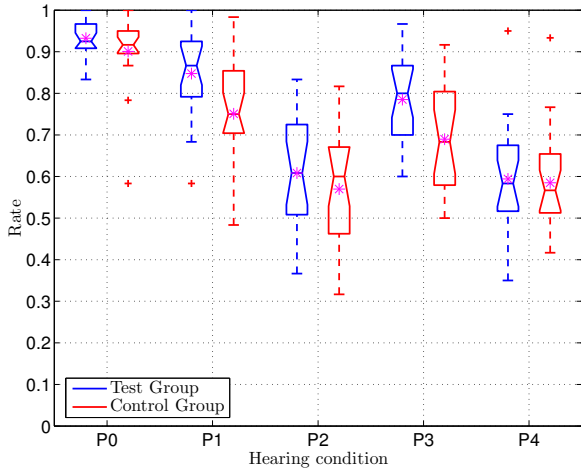


Figure 2.23: Box plot with mean values (asterisks) of the correct quadrant rate for both groups.

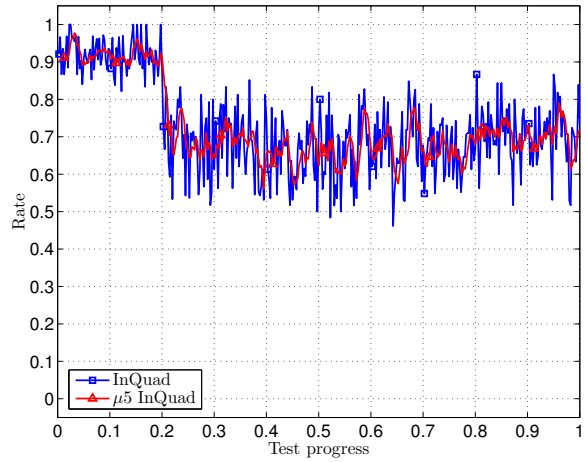


Figure 2.24: Correct quadrant rate as a function of the test progress.

localization performance. When wearing earplugs, subjects who performed training localized sound sources in the correct quadrant more often than subjects who did not perform training (p-value: < 0.05).

In the first fifth of the test, i.e. during natural hearing, the average correct response rate is 92.41 % and drops to 67.69 % for the remaining samples, i.e. during non-natural hearing. This decrease is in coherence with the previous findings that HPDs reduces the localization performance.

2.3.3.5 Confusion rate

The localization confusions are split in three types: front-back confusion, left-right confusion and mixed confusions. In Figure 2.25 the distributions and mean values of these confusion rates are visualized.

2.3.3.5.1 Front-back confusion Front-back confusions occur when the responded angle ϕ_r is in the frontal or back hemisphere, respectively, while the tested angle ϕ_t is in the other hemisphere. Sounds which are positioned on the interaural axis, i.e. $\phi_t \in \{90^\circ, 270^\circ\}$ cannot be assigned unambiguously to either the frontal or back hemisphere. Hence, we exclude the data of the calculation of the front-back confusion rate where the sound source is located on the borders of the hemisphere, i.e. $\phi_t \in \{90^\circ, 270^\circ\}$.

The lowest front-back confusion is obtained under the natural hearing, regardless of whether sub-

2.3. LOCALIZATION TEST

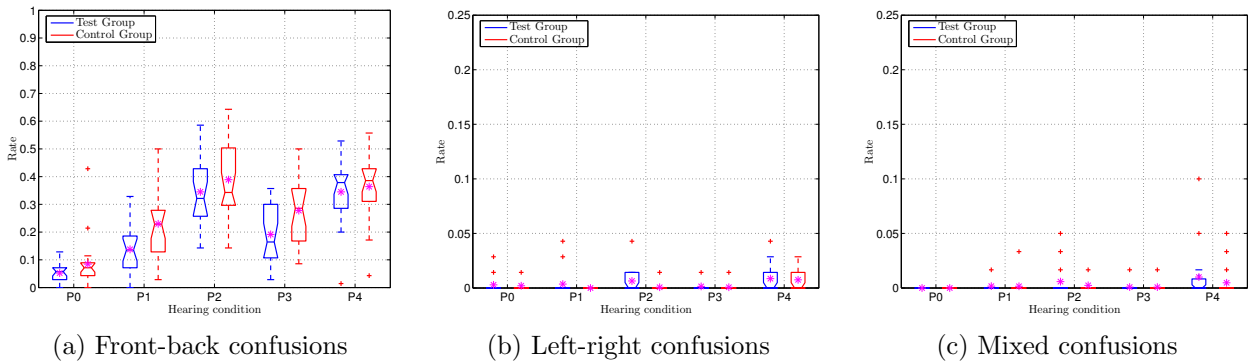


Figure 2.25: Box plots with mean values (asterisks) of the confusion rates for both groups as functions of the hearing condition. Note the different scaling of ordinate!

jects are trained or not, c.f. Figure 2.25a. Trained subjects achieve an average front-back confusion rate of 5.21 % (test group), while untrained subjects rank at 8.44 % (control group). No modifications in the HRTFs of the subjects were introduced, so this good performance is evident. Further, the localization performance with the earplugs P1 (13.79 %, 22.99 %) and P3 (19.14 %, 27.76 %) is better than with the earmuffs P4 (34.50 %, 36.39 %) and P2 (34.50 %, 38.91 %), in the test and control group respectively. The obtained front-back confusion rate with passive earmuffs match well with the value of 33 % from [182]. The effect of the HPDs on the front-back confusion rate is significant in both groups (p-values: < 0.001) except for the pairs of earplugs, i.e. between P1 and P3, and earmuffs, i.e. between P2 and P4. Further, the difference in the front-back confusion rates between HPD P2 and P3 as well as between HPD P3 and P4 is less significant in the control group (p-values: < 0.05) than in the test group (p-values: < 0.001).

The interquartile range of HPD P0 is 4.32 % (control group) and 4.60 % (test group). This interval is largely increased with non-natural hearing conditions up to 18.93 % (control group) and 20.79 % (test group). The increase of the interquartile range signify an increased spread in the data. This means that, on the one hand, HPDs seem to induce different modifications in the HRTF among the subjects, and, on the other hand, subjects are individually sensitive to modifications in their HRTFs.

It is obtained that training reduces the front-back confusion rate under all hearing conditions. The reduction is statistically significant in the case of HPD P1 (p-value: < 0.01) and P3 (p-value: < 0.05), i.e. the group of earplug HPD, but not in the case of natural hearing (p-value: $\gg 0.05$) nor when using earmuff HPDs (p-values: $\gg 0.05$).

2.3.3.5.2 Left-right confusion Left-right confusions occur when the responded angle ϕ_r is in the left or right hemisphere, respectively, while the tested angle ϕ_t is in the other hemisphere. Similar to the front-back confusion rate, data where the tested angle ϕ_t is located on the borders of the hemispheres, i.e. $\phi_t \in \{0^\circ, 180^\circ\}$, is excluded from the calculation of the left-right confusion rate.

The left-right confusion rate, as illustrated in Figure 2.25b, reaches its maximum of 0.86 % (test group) and 0.75 % (control group) with HPD P4. No left-right confusions are obtained in the control group with HPD P1. The left-right confusion rate is 0.29 % (test group) and 0.2 % (control group) with natural hearing (HPD P0). Better performance is obtained with HPD P3 in both groups and with HPD P2 in the control group only. Worse performance is obtained in the test group with HPD P2 and P3 and in both groups with HPD P4. The effect of the training is statistically significant for HPD P2 (p-value: $0.025 < 0.05$), but not for the remaining HPDs (p-value: $\gg 0.05$). The effect of the hearing condition on the left-right confusion is only significant between pairs of hearing conditions that include HPD P4 (p-values: < 0.05): The pair HPD P3–P4 in the test group; The pairs of a non-natural hearing condition and HPD P4 in the control group; The pair of the natural hearing condition and HPD P4 in the control group.

In both groups most left-right confusions are obtained with HPD P4 and, as mentioned above, it is the only hearing condition that shows an effect on the left-right confusions. Additionally, HPD P4 is a passive model while all other HPDs are active models. Thus HPD P4 corrupts not only the spectral cues but also the ITD and ILD. This prevents the subjects from left-right discrimination and leads to the increased left-right confusion rates of 0.86 % and 0.75 %.

2.3.3.5.3 Mixed confusion Mixed confusions are equivalent to simultaneous front-back and left-right confusions. They occur when the responded angle ϕ_r is located in the diagonal opposing quadrant of the tested angle ϕ_t . As for the front-back and left-right confusion, data where the tested angle ϕ_t is located on the border of the quadrant, i.e. $\phi_t \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, is excluded from the calculation of the mixed confusion rate.

With HPD P0, no mixed confusions occur in either group, c.f. Figure 2.25c. The average confusion rate is in both groups inferior to 0.1 % for HPD P3 and around 0.17 % for HPD P1. Under HPD P2, resp. P4 the mixed confusion rate in the test group (0.58 %, resp. 1 %) is about twice the mixed confusion rate in the control group (0.24 %, resp. 0.48 %). In none of the groups any significant effect

2.4. CONCLUSION

of the HPD on the confusion rate is obtained (p-values: > 0.05). Same holds for the effect of the training on the confusion rate for the entire set of tested HPDs (p-values: $\gg 0.05$).

2.3.3.6 Front-back confusion rate & hearing level

In Section 1.2.2.4 we have seen that certain frequency bands are dominant when discriminating between frontal and back sound incidence. Since the hearing thresholds varies from subject to subject, it is of interest to determine whether there is a relationship between the frequency dependent hearing threshold and the localization performance.

The influence of the hearing threshold on the localization performance, in particular the front-back confusion rate, is shown in Figure 2.26. Only frequencies above 500 Hz are shown as the aging induced shift in hearing threshold is more important for high frequencies than for low frequencies [80]. In addition, the majority of Blauert's directional and boosted bands are located above 500 Hz. Pearson's Correlation Coefficient (PCC) between the front-back confusion rate and the frequency dependent hearing threshold is -0.1 (1 kHz), -0.19 (2 kHz), -0.21 (4 kHz), -0.24 (8 kHz). For the remaining, not shown frequencies the PCCs are -0.06 (125 Hz), -0.02 (250 Hz) and -0.15 (500 Hz). For all frequencies the observed PCC are negative and close to zero, hence there is no correlation between the hearing threshold and front-back confusion rate.

2.4 Conclusion

HPD induced modifications of the HRTF and the sound localization performance with and without HPDs were examined in this chapter. The effect of HPDs on the HRTF was analyzed in an anechoic chamber on three acoustic dummy head configurations with four HPDs, including earplug, earmuff, active, and passive HPDs. Earmuff HPDs lead to highly modified HRTFs, while earplug HPDs introduce much less modifications. The bandwidth of the electroacoustic systems of active HPDs is important to provide sufficient high frequency spectral cues. The effect of the type of earplug HPDs was evaluated. Universal and custom molded earplugs introduce modifications on the HRTF, which do not show noticeable differences.

A VAE has been designed and set up in a semi-anechoic chamber in order to conduct sound localization tests. With this VAE the sound localization performance of 40 subjects, with and without

2.4. CONCLUSION

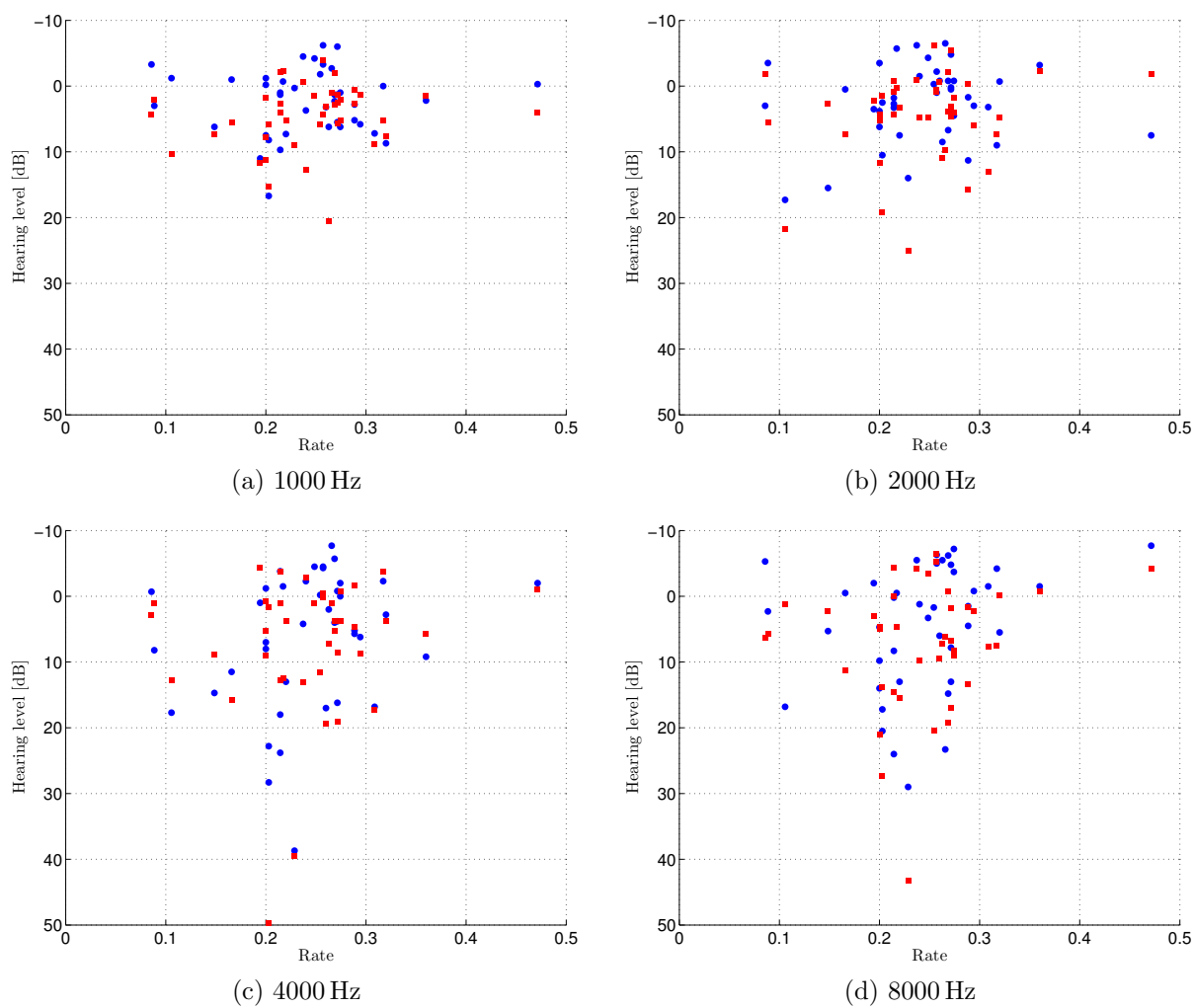


Figure 2.26: Hearing level at the left (dot), resp. right (square) ear confronted with the front-back confusion rate for frequencies ≥ 1 kHz.

2.4. CONCLUSION

HPDs, was assessed. Subjects achieve best localization performance with their natural hearing. The probability of obtaining a correct response without HPDs is at least by 14 % higher than with HPD. It is obtained, that subjects respond either correctly or with a front-back confusions. The number of obtained left-right and mixed confusions are negligible. The decrease of sound localization performance when wearing HPDs is due to the increase of front-back confusions. Earplug HPDs allow better sound localization than earmuff HPDs. An increased number of correct responses and a reduced number of front-back, left-right, and mixed confusions are obtained when using earplug HPDs rather than earmuff HPDs. Hence, earplug HPDs are advantageous over earmuff HPDs concerning sound localization performance. Small differences were obtained between the earplug HPD models, which are due to the size of the earplug and consequently the amount of the covered outer ear [164]. HPD P3 covers the cavum concha and cymba concha while HPD P1 covers only the cavum concha, see [61] for definitions. The height of the cavum concha, resp. cymba concha are the most, resp. third most important feature of the outer ear, influencing the HRTF and sound localization [184].

A strong interaction was identified between the objectively measured RMSE of HRTF induced by HPDs and the subjective localization performance with each of the HPD. Figure 2.27 visualizes this effect by confronting the correct response rate (averaged over both groups) under the tight condition ($\phi_i = 22.5^\circ$) with the RMSE (frequency range “speaker”) for each of the non-natural hearing conditions. A PCC of -0.96 supports that the more an HRTF is modified the more the localization performance drops. Further, front-back confusions occurred during the localization test only when subjects wore HPD, hence HPDs modify mainly the spectral cues which are important for front-back discrimination. HPD P1 and P3 introduce less modifications in the HRTF and lead to enhanced localization performance than HPD P2 and P4. Consequently, it is preferable that HPDs be designed as earplug models, with the earplugs being as small as possible to cover as little area of the outer ear as possible.

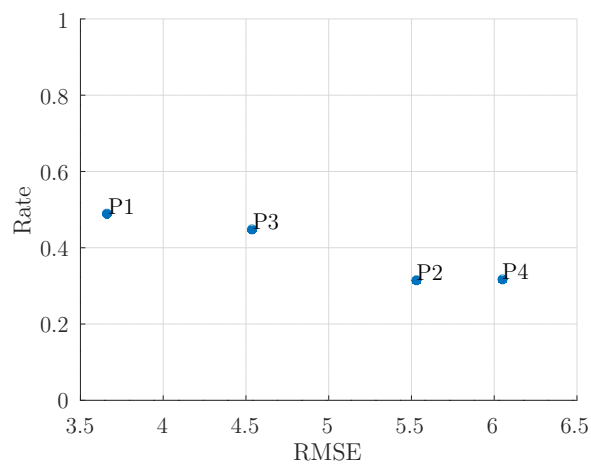


Figure 2.27: Relation between the correct response rate (tight condition, $\phi_i = 22.5^\circ$) and RMSE (x-axis) for the non-natural hearing conditions HPD P1 to P4.

2.4. CONCLUSION

Chapter 3

Investigations on the HRTF and its components

Content

3.1	HRTF simulation	100
3.2	Contributions of the outer ear	103
3.3	Recombination of the HRTF	105
3.4	Conclusion	107

In the previous chapter it was obtained, that earplugs are advantageous over earmuffs regarding localization performance. In order to show the importance of working on the development of earmuffs rather than earplugs, the current chapter focuses on the contributions of the outer ear on the natural HRTF and directs the following work towards HPDs of type earmuff.

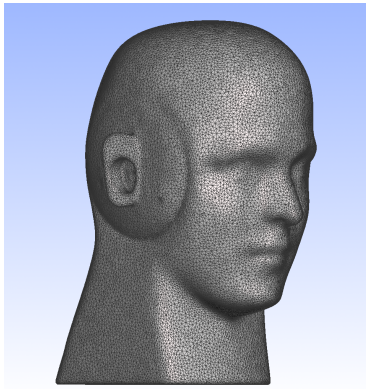
3.1 HRTF simulation

The natural HRTFs of dummy head Harry34, Harry33, and HATS33 were numerically simulated. For this, the dummy heads were scanned to obtain their geometrical 3D representation, c.f. Section 1.4.1.2. To reduce the size of the 3D models, the base plates of Harry34 and Harry33 and the torso of HATS33 were removed. The final 3D models are shown in Figures 3.1a, 3.1c and 3.1e, counting 69985 (Harry34), 62686 (Harry33) and 46666 (HATS33) triangular surface elements. The HRTFs were simulated for the same frequency scale and angles of incidence which were used for the HRTF measurements, i.e. 1600 logarithmically spaced frequencies in the interval [16 Hz; 25.6 kHz] and 16 angles of incidence spaced by 22.5° in the interval $[0^\circ; 360^\circ]$. The magnitudes of the simulated HRTFs are shown in Figures 3.1b, 3.1d and 3.1f.

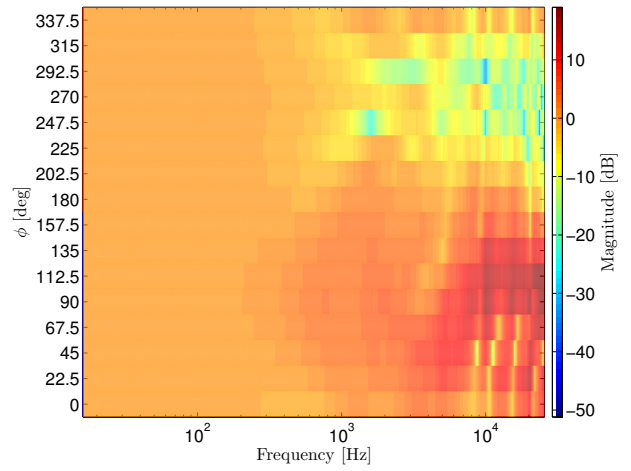
The general characteristics and in particular the locations of the peaks and notches in the simulated HRTFs, c.f. Figure 3.1, and the measured HRTFs, c.f. Figures 2.5a, 2.6a and 2.7a, agree with each other and with Mokhtari's observations [185]. Particular attention must be paid to the simulated HRTF of HATS33 in Figure 3.1f where low frequency ($f < 30$ Hz) details occur for all angles of incidence. They do not originate from the HRTF, rather they are numerical artifacts. The underlying Conjugate-Gradient Solver (CGS) of the simulation does not converge within the maximum number of 1500 iterations [159]. Fixing this problem does not seem reasonable, since the HRTF measurement does not provide reliable data for frequencies below 30 Hz and thus a comparison between measurement and simulation below 30 Hz is not possible anyway. Above 30 Hz the CGS converges within 41 to 307 iterations (average: 54.33, standard deviation: 15.97).

The RMSEs between the measured and simulated HRTFs was calculated based on the RMSE definition in Equations (2.1) and (2.2). The equations were adapted and we replaced HRTF_{P_i} , resp. HRTF_{P_0} by the measured, resp. simulated natural HRTF. The RMSE between simulation and measurement, calculated on the entire frequency range, equals 2.88 dB (Harry34), 2.58 dB (Harry33), and

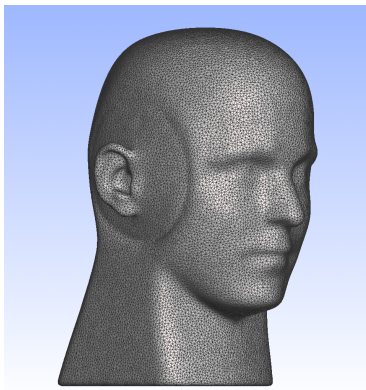
3.1. HRTF SIMULATION



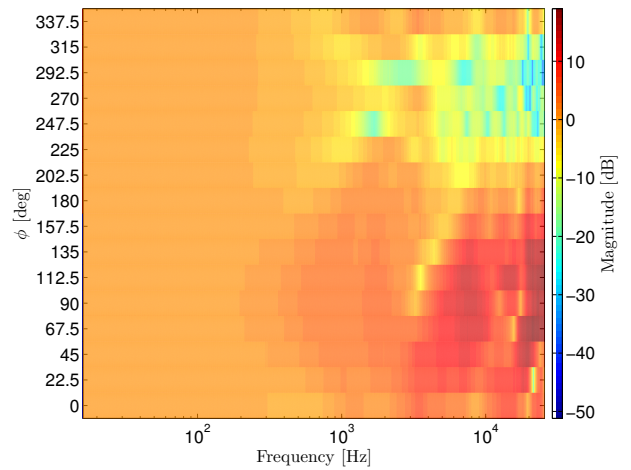
(a) Harry34



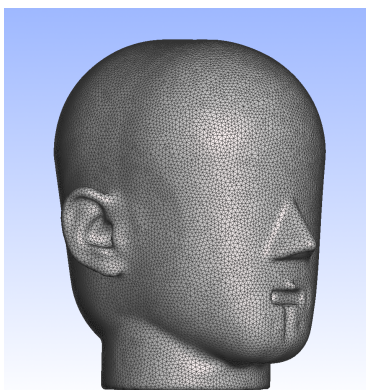
(b) HRTF of Harry34



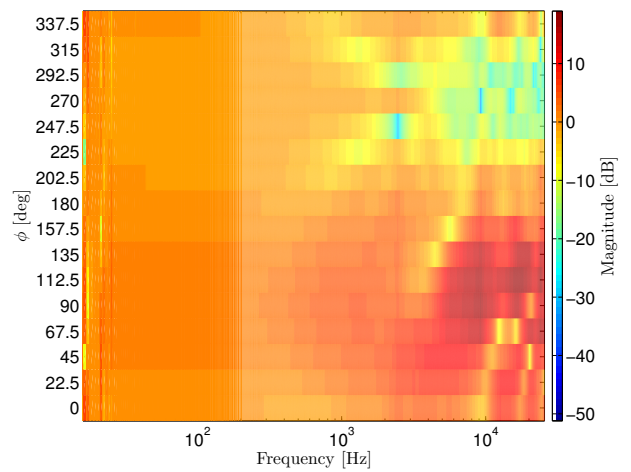
(c) Harry33.



(d) HRTF of Harry33



(e) HATS33.



(f) HRTF of HATS33

Figure 3.1: Left: 3D models of dummy heads. Right: Magnitude of numerically simulated HRTFs.

3.1. HRTF SIMULATION

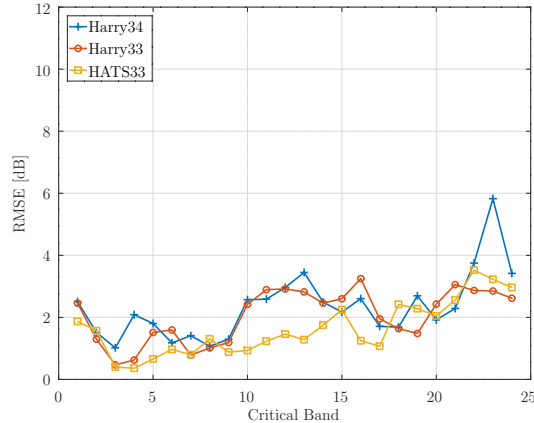


Figure 3.2: RMSE in the critical bands [12] between the simulated and measured HRTFs for the three dummy heads.

2.19 dB (HATS33), while considering the loudspeaker’s frequency range, it equals 2.54 dB (Harry34), 2.10 dB (Harry33), and 1.91 dB (HATS33). In comparison, the RMSE between pairs of natural HRTFs of different dummy heads varies between 2.58 dB and 2.97 dB (intra measurement), 2.22 dB and 2.84 dB (intra stimulation), 2.55 dB and 3.07 dB (inter measurement-stimulation), regarding the frequency range of the loudspeaker. With the exception of Harry34, the RMSE between measured and simulated HRTFs of the same dummy head configuration is lower than the RMSE between pairs of HRTFs of different dummy head configurations. The poor repeatability of the measurements, in particular with Harry34, causes the increased RMSE between measurement and simulation. Mokhtari et al. obtained a spectral distance between simulation and measurement varying between 3.8 dB to 4.7 dB [185]. They face the problem of positional errors as working with human subjects, who slightly move during the HRTF measurement and 3D scanning procedure. Further, they used different head-to-source distances for the simulation and measurement. In contrast, for the simulations and measurements, we used dummy heads that move much less than human subjects and used equal head-to-source distances.

Figure 3.2 visualizes the RMSE between the simulated and measured HRTFs per critical band. Across the entire frequency range, the RMSE ranges the interval [0.36 dB; 5.83 dB]. Between critical band N° 10 and N° 15, i.e. [1.08 kHz; 2.7 kHz], the RMSEs are smaller for HATS33 than for Harry34 and Harry33. It is identified that at contralateral directions with $\phi \in [247.5^\circ; 292.5^\circ]$ the difference between the simulated and measured HRTF of HATS33 is noticeably lower than for Harry34 and Harry33. Small misalignment of the dummy head axis with respect to the reference axis during the

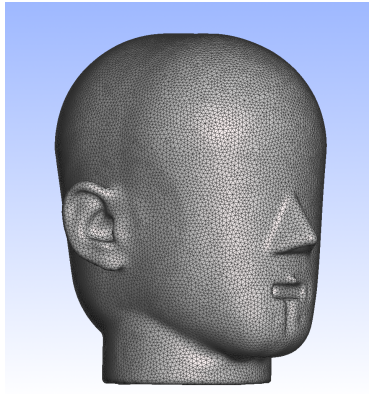
measurements and in the 3D models are subjected to cause these effects. Further, it is noticed that the RMSE tends to increase with increasing frequency. The 3D scanning process captured the geometry of the dummy heads with a spatial resolution of 1 mm [158]. The reshaping of the meshes to an edge length of 2.23 mm reduced their number of elements. Hence, there are fine geometrical deviations between the real-life dummy heads and their numeric siblings, i.e. 3D meshes. These deviations become more and more noticeable in the high frequency range of the HRTFs which is in accordance with the observations of Braren et al. [186].

3.2 Contributions of the outer ear

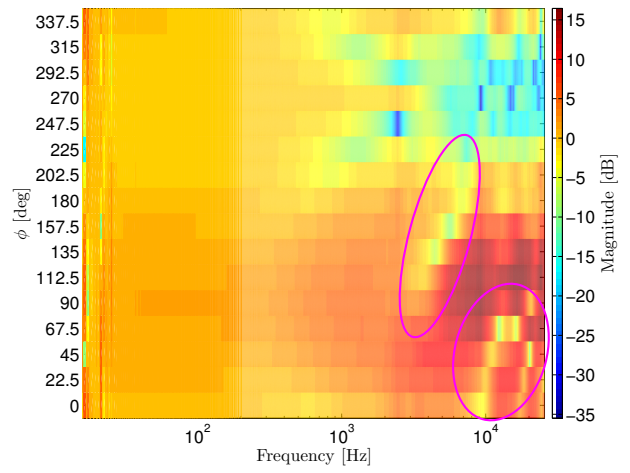
Despite its name Head-Related Transfer Function the HRTF is defined by the geometries of the ensemble of outer ears, head, and shoulders, c.f. Section 1.2.2.5. Like the HRTF, the Pinna-Related Transfer Function (PRTF) describes the directional and frequency dependent transfer function but taking only the geometry of the outer ear into account. The PRTF contains major spectral information above 3 kHz due to the fine geometrical structure of the outer ear [187]. Shaw, Takemoto et al., and Otani et al. identified resonance frequencies of the outer ear between 3.5 kHz and 17 kHz which they were able to relate to outer ear geometries, such as concha height, width, and depth [188, 153, 151]. In the following, spectral cues related to the PRTF are confronted with spectral cues related to the HRTF.

To compare PRTF and HRTF, the 3D models of a new dummy head, called “HATS00”, and of the outer ears were constructed. HATS00 was obtained by removing the outer ears from HATS33’s 3D model and adding a left and right ear canal, each measuring 15 mm in depth and 8 mm in diameter, c.f. Figure 3.3c [189]. The depth corresponds to the depth of the ear canal in the Harry-based dummy heads. The 3D models of the removed outer ears were individually centered at the origin and attached to a circular support. The resulting model of the right outer ear is shown in Figure 3.3e. Numerical simulations were performed with the left and right outer ears and HATS00, using the same simulation settings as in Section 3.1. These 3D models counted 44514 (HATS00) and 4854 (outer ear) elements. Compared with the 3D model of HATS33, the number of elements was reduced by 5% (HATS00) and 90% (outer ear). The magnitude of the HRTFs of HATS33 and HATS00 and the PRTF are shown in Figures 3.3b, 3.3d and 3.3f. For HATS00, c.f. Figure 3.3d, it can be noticed that the frequency dependent diffraction starts by 300 Hz at $\phi = 90^\circ$. Additionally, the “bright spot

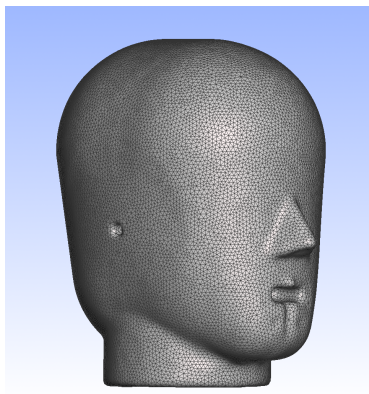
3.2. CONTRIBUTIONS OF THE OUTER EAR



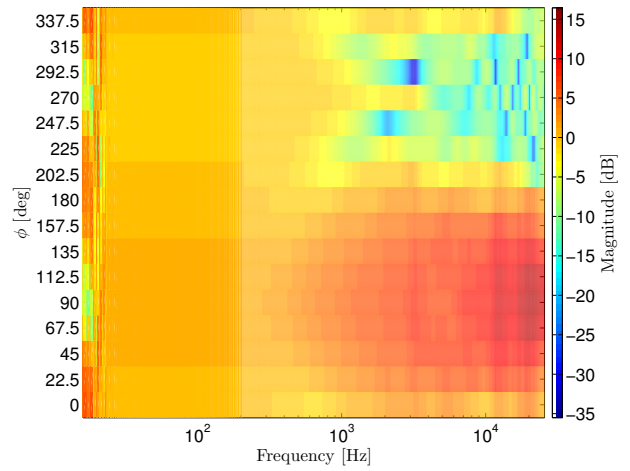
(a) HATS33.



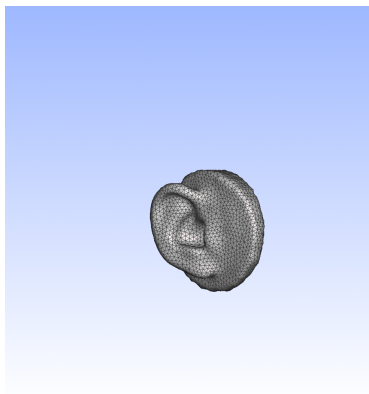
(b) HRTF of HATS33



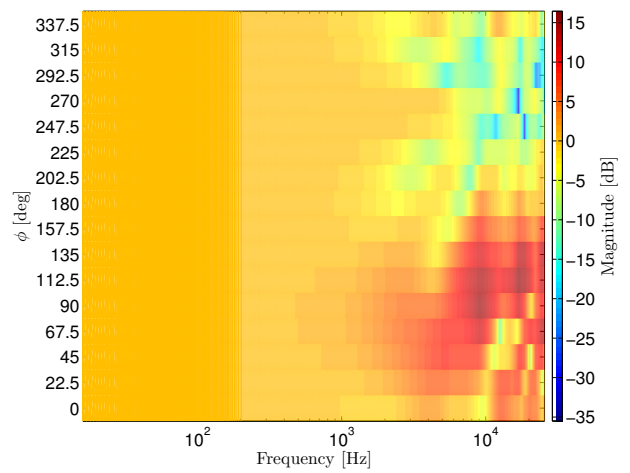
(c) HATS00.



(d) HRTF of HATS00.



(e) HATS33's ear simulator.



(f) PRTF of HATS33's ear simulator.

Figure 3.3: Left: 3D model of HATS33 and the deviation models HATS00 and HATS33's ear simulator, all shown with equal scale. Right: Numerically simulated HRTFs and PRTF.

effect” at $\phi = 270^\circ$ is most distinguishable around 2 kHz, aligning with the characteristics of the ILD, c.f. Figure 1.7. By visual comparison of Figures 3.3b and 3.3d it turns out that HATS00, i.e. the geometry of the head, defines the basic structure of the HRTF, including ILD and ITD. The HRTF of HATS00 is highly symmetric to $\phi = 90^\circ$ particularly for ipsilateral positions, leading to equal spectral cues for directions of incidence symmetric to the interaural axis. Front-back discrimination is not possible with such symmetric cues as they introduce equal directional information for frontal and back sounds. Looking at the PRTF in Figure 3.3f, non-symmetric spectral cues are obtained above 1 kHz, providing unambiguous directional information. The geometries of the outer ear are interacting with high frequencies only. Two important notches are identified: The first propagates from 10 kHz and 0° to 12 kHz and 67.5° , the second propagates from 6 kHz and 112.5° to 10 kHz and 247.5° . These notches can be easily recognized in the HRTF of HATS33, c.f. magenta highlighted regions in Figure 3.3b. As noted by Algazi et al., Takemoto et al., and Otani et al., the HRTF can be decomposed into spectral cues contributed by the head on the one hand and the outer ear, i.e. the PRTF, on the other hand [187, 153, 151]. Only the combination of the head together with the outer ears leads to spectral cues that are unique for each direction of incidence and thus allow reliable sound location with natural hearing. The head without outer ears does not allow front-back discrimination, since it introduces symmetric cues.

3.3 Recombination of the HRTF

The recombination of a generic HRTF, e.g. of a sphere or a head without ears, in the following called “sHRTF”, with the individual PRTFs of subjects can reduce the effort required to obtain individual HRTFs for a large number of subjects [187]. The basic shape of the HRTFs is defined by the sHRTF, whose simulation is expensive regarding computational resources and time due to the size of the 3D mesh. In contrast, this simulation must be performed only once as the sHRTF is used for all subjects. The simulations of the PRTFs with less complex models are cheap and fast, but must be performed individually for each subject. The resulting individual PRTF contains individual spectral cues which are essential for front-back discrimination and elevation determination. Algazi et al. propose to combine the sHRTF and PRTF by summation of their logarithmic magnitudes, i.e. in the dB domain [187]. We tested two methods of combining the sHRTF and PRTF: summation and arithmetic averaging. To better deal with the phase and gain of the transfer function, we combined the

3.3. RECOMBINATION OF THE HRTF

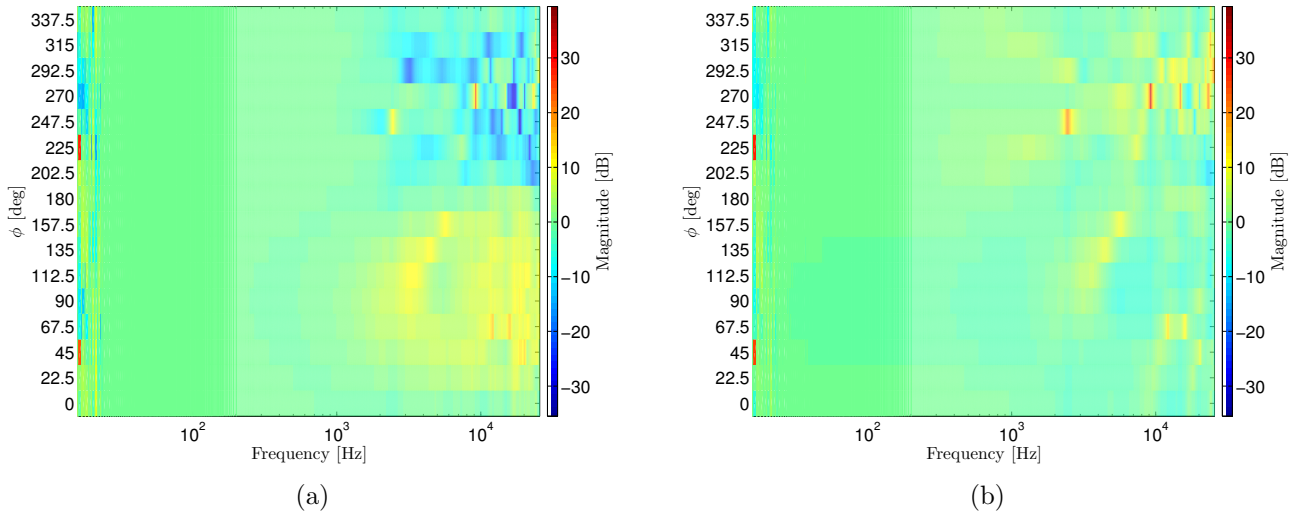


Figure 3.4: Error function between the HRTF of HATS33 and the composed HRTFs, using the sum method (left) and using the mean method (right). Positive values indicate an overestimation of the original HRTF.

previously simulated sHRTF and PRTF in the complex domain. This required to use multiplication instead of summation the geometric average instead of the arithmetic average.

The two recombination approaches are compared by evaluating the error function between their results and the originally simulated HRTF of HATS33. The error functions are plotted in Figure 3.4. Independently of the method, small errors are obtained between 50 Hz and 1 kHz. Above 1 kHz the error functions are different. In general, the sum method leads to overestimation at ipsilateral positions and underestimation at contralateral positions, while the mean method leads to overestimation on the entire horizontal plane. The RMSEs between the recombined and original HRTFs was calculated with Equations (2.1) and (2.2) and replacing HRTF_{P_i} , resp. HRTF_{P_0} by the recombined, resp. original HRTFs. For the entire horizontal plane the RMSEs are 4.21 dB (sum) and 2.41 dB (mean) in the frequency interval [100 Hz; 18 kHz]. Considering the ipsilateral, resp. contralateral positions, the RMSEs are 4.04 dB (sum) and 2.20 dB (mean), resp. 4.87 dB (sum) and 2.80 dB (mean). We notice that these values differ by approximately a factor of 2. This is related to the similarity of the sum and mean method, which likewise differ in a factor of 2. The RMSEs does not consider the relative magnitude of local extreme values in the HRTF. Even though the mean method results in lower RMSE, the averaging damps the depth of notches and the height of peaks, which are important as they lead to individually shaped HRTFs. At this point, further investigations involving subjective localization

tests with the original and recombined HRTFs should be performed with human subjects .

3.4 Conclusion

It turns out that the PRTF and thus the outer ear contributes most of the individual, directional, spectral cues of the HRTF. Concerning front-back discrimination and elevation determination the PRTF plays an essential role for sound localization. Conversely, this means that HPDs should cover the outer ear the least possible. In the case of earplug HPDs, this requires miniaturized earplugs, as they have been developed and are applied for hearing aids. Hearing aids are available as in-the-ear (placed in the concha), in-the-canal (placed at the ear canal entrance), and completely-in-the-canal (placed in the ear canal) devices [190, 191]. The more the receiver of the hearing aid is in the ear canal, the least modifications are induced in the HRTF, and the fewest sound localization errors are obtained [192, 193, 169]. These miniaturized designs should be considered when improving sound localization in the context of earplug HPDs. Nevertheless, the process of miniaturization must not influence with the primarily purpose of providing acoustic protection.

In contrast, earmuff HPDs inherently enclose the outer ear, making the miniaturization of their geometries, as proposed for earplug HPDs, impossible. At the same time, the large size of earmuffs allows an easy and fast handling, even wearing gloves, neither have large sized devices to be inserted by specialist unlike completely-in-the-canal devices [191]. Therefore, earmuffs advantages over earplugs and are therefore not completely replaceable by them. In order to augment sound localization performance with earmuffs, other techniques than reduction of the geometry must be envisaged. In the following, different designs of earmuff HPDs and signal processing methods are studied, with respect to sound localization in the horizontal plane.

3.4. CONCLUSION

Chapter 4

Spatial audio filters

Content

4.1	Parametric Difference-HRTF	110
4.1.1	Front-back differences in HRTF	110
4.1.2	Parameterized filters	111
4.2	Listening Test	120
4.2.1	Setup	120
4.2.2	Stimuli	120
4.2.3	Procedure	128
4.2.4	Participants	130
4.2.5	Results	130
4.2.6	Discussion	142
4.3	Conclusion	147

In order to reconstruct the directional information of environmental sounds under earmuffs, we analyze in this chapter the spectral difference in HRTFs between frontal and back sound incidence. Further, a boost band filter as function of the azimuth angle ϕ , approximating the identified spectral differences, is developed. The developed filter is evaluated against filters from literature regarding localization performance. For this a headphone based VAE is rendered and a listening test is conducted.

4.1 Parametric Difference-HRTF

Filters that introduce spectral cues have been either detailed HRTF-based filters as functions of the azimuth and elevation angle, or equalizer-type boost or damping filters that distinguish only between frontal and back sound incidence. Our aim was to develop a filter with a hybrid structure based on the spectral front-back difference of HRTFs by combining equalizer-like band filters with the azimuth dependent characteristics of an entire HRTF.

4.1.1 Front-back differences in HRTF

The spectral front-back difference HRTF, called Δ HRTF, was defined on pairs of azimuth angles which are symmetrically positioned around the interaural axis, c.f. Equation (4.1). Equivalent to the HRTF, the Δ HRTF is a function of the frequency f and azimuth angle ϕ .

$$\Delta\text{HRTF}(f, \phi) = \text{HRTF}(f, \phi) - \text{HRTF}(f, \alpha - \phi) \quad \begin{cases} \alpha = 180^\circ & \text{for } \phi \leq 180^\circ \\ \alpha = 540^\circ & \text{for } \phi > 180^\circ \end{cases} \quad (4.1)$$

We focused on the ipsilateral ear, i.e. the HRTF data for azimuth angles between 0° and 180° . The Δ HRTF was calculated for the measured and simulated, natural HRTF of the three dummy heads Harry33, Harry34, and HATS33.

Figure 4.1 shows the six Δ HRTF for the four pairs of azimuth angles $0^\circ/180^\circ$ (referred to as 0°), $22.5^\circ/157.5^\circ$ (referred to as 22.5°), $45^\circ/135^\circ$ (referred to as 45°), and $67.5^\circ/112.5^\circ$ (referred to as 67.5°). Apart from the effects due to the limited bandwidth of the loudspeaker and the imperfectly compensated impedance peaks of the loudspeaker, the amplitudes are close to 0 dB below 1 kHz. Hence, no spectral difference is noticed in this frequency range. Between 1 kHz and 10 kHz common characteristics are identified. A broad boost band, c.f. blue bordered region, is superposed with a narrow peak, c.f. green bordered region, and followed by a small damping band in the high frequencies.

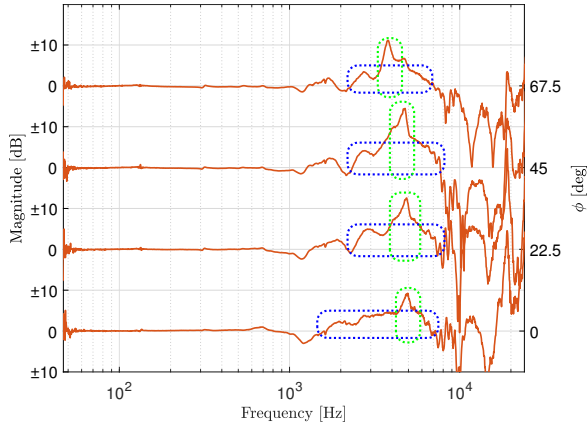
The broad boost band is defined by the broadest, coherent frequency range where $\Delta\text{HRTF} > 0$ dB. The width of the narrow peak is defined by the 3 dB bandwidth, centered around the maximum which is in the broad boost band. The broad boost band, the narrow peak, and the damping band alter their positions on the frequency axis and magnitudes along with ϕ . The broad boost bands, resp. small damping bands are mainly located in the frequency interval [2 kHz; 8 kHz], resp. [8 kHz; 10 kHz]. The location of the narrow peak varies in the interval [3 kHz; 7 kHz]. The intervals of the broad boost bands, resp. the positions of the narrow peaks correspond well with Blauert's boosted, resp. directional bands for frontal sound incidence which range [1.86 kHz; 7.03 kHz], resp. [2.91 kHz; 5.65 kHz], c.f. Table 1.1 and [40].

The broad boost band shifts to lower frequencies with increasing azimuth angle and in amplitude and bandwidth. The narrow peak is wide in bandwidth and small in amplitude for median positions ($\phi = 0^\circ$) and sharp, i.e. narrow, in bandwidth and large in amplitude for lateral positions ($\phi = 67.5^\circ$). For $\phi = 45^\circ$ the peak reaches its maximum. Due to the definition in Equation (4.1) the magnitude of the ΔHRTF collapses towards a zero line at $\phi = 0^\circ$.

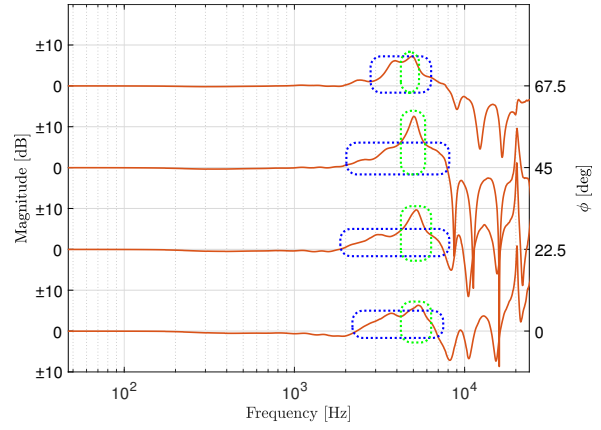
4.1.2 Parameterized filters

The parameterized spectral front-back difference HRTF, called $\hat{\Delta}\text{HRTF}$, was developed, based on the measured and simulated HRTF of dummy head Harry33. By this we considered the advanced, human-like outer-ear simulators and avoided the highly narrow band peaks in the ΔHRTF of HATS33 which are poorly approximated by low order, digital filters. The magnitude of the $\hat{\Delta}\text{HRTF}$ is defined by the magnitude of the ΔHRTF of dummy head Harry33 at pairs of $0^\circ/180^\circ$, $22.5^\circ/157.5^\circ$, $45^\circ/135^\circ$, and $67.5^\circ/112.5^\circ$. Linear regression was applied on these data points to obtain analytic functions and interpolate the magnitude of the ΔHRTF at intermediate angles. $\hat{\Delta}\text{HRTF}$ is a function of the frequency f and the azimuth angle ϕ . The magnitude of the $\hat{\Delta}\text{HRTF}$ is inverse symmetrical around the interaural axis. Hence, the corresponding filters must have half the amplitude of ΔHRTF . The implementation of $\hat{\Delta}\text{HRTF}$ is based on linear combination of 2nd-order Infinite Impulse Response (IIR) peak filters and their implementation proposed by Zölzer in [194]. IIR filters reproduce a given frequency response with less filter coefficients than Finite Impulse Response (FIR) filters. The resultant simplicity allows real time implementation on nearly any embedded systems with little computation power [195].

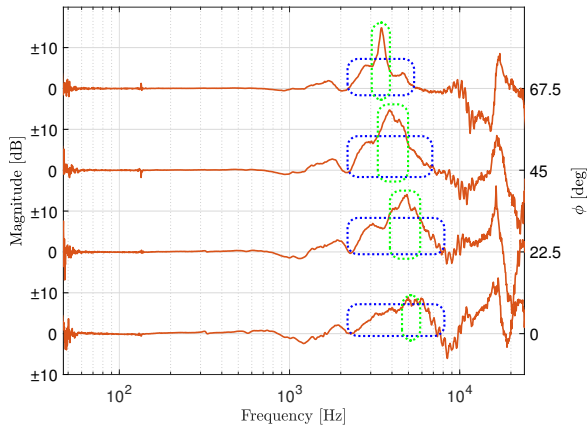
4.1. PARAMETRIC DIFFERENCE-HRTF



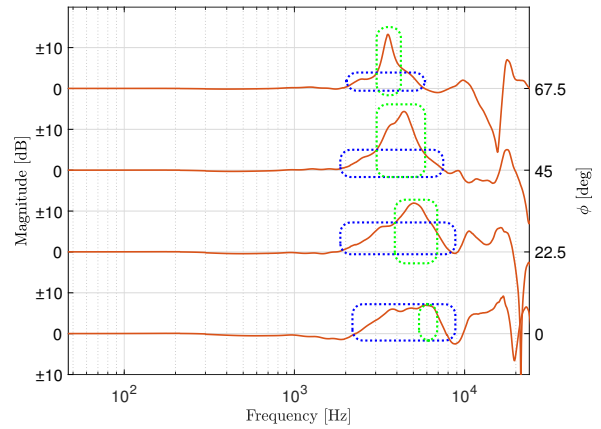
(a) Harry34 measured.



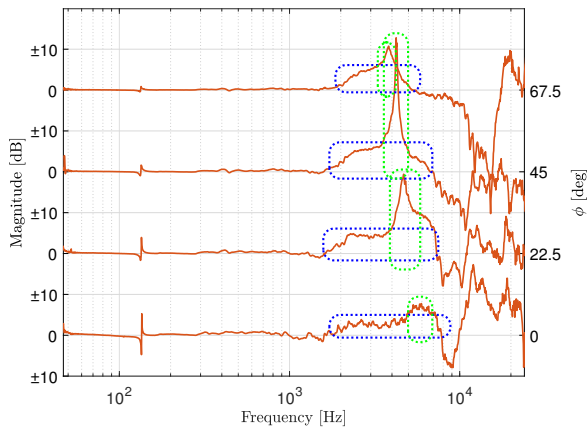
(b) Harry34 simulated.



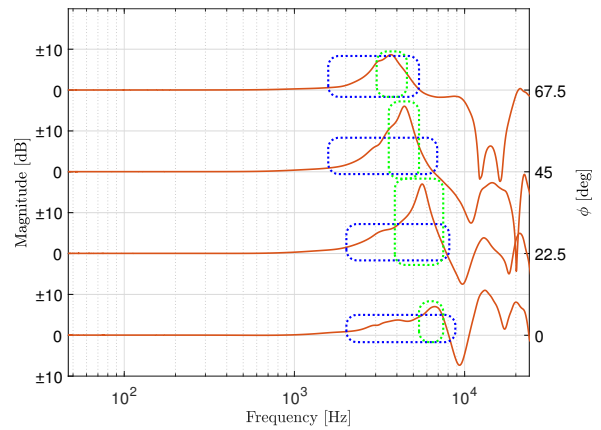
(c) Harry33 measured.



(d) Harry33 simulated.



(e) HATS33 measured.



(f) HATS33 simulated.

Figure 4.1: Spectral front-back difference (Δ HRTF) for the measured (Figures 4.1a, 4.1c and 4.1e) and the simulated (Figures 4.1b, 4.1d and 4.1f) HRTFs of the dummy heads. The broad boost bands, resp. narrow peaks are bordered by blue, resp. green rectangles.

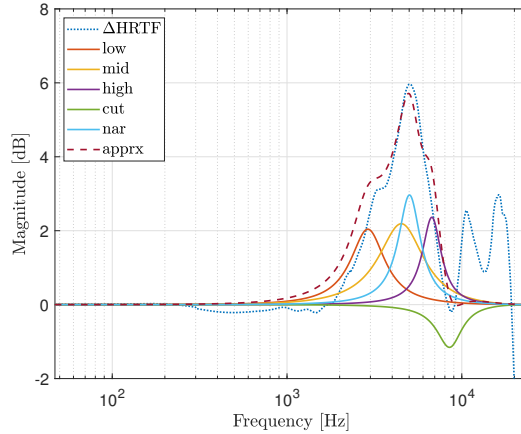


Figure 4.2: Example of approximating the front-back difference HRTF (Δ HRTF) at the azimuth pair $22.5^\circ/157.5^\circ$. The 2nd-order IIR filters “low”, “mid”, “high”, “cut”, and “nar” are linearly combined and result in the filter “apprx”. Here, the RMSE between the filter “apprx” and Δ HRTF is 1.75 dB.

4.1.2.1 Basic design

The approximation of the Δ HRTF is exemplary shown in Figure 4.2 for the simulated HRTF of Harry33 at the azimuth pair $22.5^\circ/157.5^\circ$. The magnitude of the original Δ HRTF is approximated by the magnitude of the frequency response of filter “apprx”. Nowak et al. proposed an automatic technique to approximate HRTFs. They required between 3 and 30 peak filters to respect the maximum error of 2 dB, but most often 7 to 8 peak filters were sufficient [75]. Our approach required a fixed number of filters, so we defined the filter “apprx” by a linear combination of five 2nd-order IIR filters. Instead of a linear combination of two 2nd-order HSF, we used three of these 2nd-order filters, called “low”, “mid”, “high”, to form the broad boost band characteristics. This increases the total filter order by two but allows in a future development stage different gains at the lower and upper edge of of the broad boost band. The steep high frequency slope of the broad boost band is obtained by the filter “cut”. The narrow peak is generated by the filter “nar”. In the following, the center frequencies and bandwidth are defined, followed by the determination of the gains.

4.1.2.2 Center frequency and bandwidth

The five peak filters are each defined by a center frequency f_0 , bandwidth f_b and gain g . In a first step, f_0 and f_b were manually determined from the Δ HRTF in Figures 4.1c and 4.1d and are listed in Table 4.1. A constant bandwidth of 4 kHz was chosen for the “cut” filter.

4.1. PARAMETRIC DIFFERENCE-HRTF

0°		“low”	“mid”	“high”	“cut”	“nar”
f_0	m	3.22	4.60	6.44	8.50	6.11
	s	3.45	4.97	6.62	8.50	5.22
f_b	m	1.40	3.38	1.21	4.00	0.43
	s	1.50	2.72	1.27	4.00	1.71

(a)

22.5°		“low”	“mid”	“high”	“cut”	“nar”
f_0	m	3.13	4.39	6.63	8.50	4.73
	s	2.90	4.51	6.74	8.50	5.02
f_b	m	1.10	3.62	1.69	4.00	0.70
	s	1.45	3.05	2.06	4.00	1.64

(b)

45°		“low”	“mid”	“high”	“cut”	“nar”
f_0	m	3.03	3.86	5.68	7.50	3.92
	s	2.69	3.89	6.12	7.50	4.33
f_b	m	0.99	2.58	2.27	4.00	0.74
	s	0.80	3.14	1.83	4.00	0.94

(c)

67.5°		“low”	“mid”	“high”	“cut”	“nar”
f_0	m	2.99	3.60	4.64	6.00	3.49
	s	2.71	3.48	4.82	6.00	3.56
f_b	m	1.10	1.81	1.59	4.00	0.19
	s	0.52	2.39	0.94	4.00	0.38

(d)

Table 4.1: Center frequencies f_0 and bandwidths f_b in kHz for the five 2nd-order IIR filters, extracted from the measurement (“m”) and simulation (“s”) based Δ HRTF of Harry33. Values are indicated for the HRTF pairs of 0°/180° (Table 4.1a), 22.5°/157.5° (Table 4.1b), 45°/135° (Table 4.1c), and 67.5°/112.5° (Table 4.1d).

For each of the five 2nd-order IIR filters “low”, “mid”, “high”, “cut”, “nar”, linear regression was used to fit 2nd-order polynomials to the data in Table 4.1. The obtained analytic functions describe the relationship between the center frequency f_0 , resp. the bandwidth f_b and the azimuth angle ϕ . For the fitting process we considered the measurement-based and simulation-based data. The obtained polynomials for the center frequencies f_0 , resp. bandwidth f_b are given in Equation (4.2), resp. Equation (4.3).

$$\mathbf{f}_0(\phi) = \begin{bmatrix} f_0^{low}(\phi) \\ f_0^{mid}(\phi) \\ f_0^{high}(\phi) \\ f_0^{cut}(\phi) \\ f_0^{nar}(\phi) \end{bmatrix} = \begin{bmatrix} 0.16 & -17.73 & 3338 \\ 0.0014 & -19.17 & 4810 \\ -0.65 & 16.69 & 6557 \\ -0.62 & 7.22 & 8538 \\ 0.092 & -38.09 & 5671 \end{bmatrix} \cdot \begin{bmatrix} \phi^2 \\ \phi \\ 1 \end{bmatrix} \quad [\text{Hz}] \quad (4.2)$$

$$\mathbf{f}_b(\phi) = \begin{bmatrix} f_b^{low}(\phi) \\ f_b^{mid}(\phi) \\ f_b^{high}(\phi) \\ f_b^{cut}(\phi) \\ f_b^{nar}(\phi) \end{bmatrix} = \begin{bmatrix} 0.043 & -13.18 & 1473 \\ -0.51 & 19.87 & 3075 \\ -0.70 & 48.69 & 1210 \\ 0 & 0 & 4000 \\ -0.33 & 10.09 & 1077 \end{bmatrix} \cdot \begin{bmatrix} \phi^2 \\ \phi \\ 1 \end{bmatrix} \quad [\text{Hz}] \quad (4.3)$$

The polynomials for the center frequencies result in R^2 values of 68.09 %, 94.55 %, 96.57 %, 97.40 %, resp. 90.78 % for the “low”, “mid”, “high”, “cut”, resp. “nar”. The polynomials for the bandwidth result in R^2 values of 65.37 %, 69.08 %, 71.29 %, N/A, and 41.80 %, respectively. The poor R^2 values are

due to the large spread of the measurement-based and simulation-based data. Due to the constant bandwidth of the “cut” filter the data has zero standard deviation. For this case the R^2 value is not defined.

4.1.2.3 Gains

The combination of the filters “low”, “mid”, “high” and “cut” is intended to shape the broad boost band of the Δ HRTF. The gain of this broad boost band (g^{bbb}) was manually determined from Figures 4.1c and 4.1d to be 5.50 dB, resp. 6.00 dB (0°), 6.50 dB, resp. 6.00 dB (22.5°), 6.65 dB, resp. 3.00 dB (45°), and 5.68 dB, resp. 2.20 dB (67.5°) for the measurement, resp. simulation based data. The narrow peak exceeds the broad boost band by 3.50 dB, resp. 1.00 dB (0°), 7.50 dB, resp. 5.93 dB (22.5°), 8.09 dB, resp. 11.36 dB (45°), and 9.17 dB, resp. 11.00 dB (67.5°) for the measured, resp. simulated HRTF data. These values define the gain g_{nar} . With linear regression we fitted 2nd-order polynomials to the gains and obtained the analytical relations between the azimuth angle ϕ and the gains as given in Equation (4.4). The constant gain of $g_{cut} = 0.1$ dB at f_0^{cut} was identified to model sufficiently the high frequency slope of the broad boost band.

$$\mathbf{g}(\phi) = \begin{bmatrix} g^{bbb}(\phi) \\ g^{cut}(\phi) \\ g^{nar}(\phi) \end{bmatrix} = \begin{bmatrix} -0.00068 & 0.016 & 5.87 \\ 0 & 0 & 0.1 \\ -0.0020 & 0.25 & 2.19 \end{bmatrix} \cdot \begin{bmatrix} \phi^2 \\ \phi \\ 1 \end{bmatrix} \quad (4.4)$$

The spread of the gain data by more than factor 2 at 45° and 67.5° resulted in the low R^2 value of 29.42 % for the broad boost band. The regression of g^{nar} represents well ($R^2 = 87.13$ %) the measured gains of the narrow peak.

4.1.2.4 Gain optimization

Regarding $g^{bbb}(\phi)$ and $g^{cut}(\phi)$ in Equation (4.4) we obtained functions which are describing the gain, resp. amplitude of the broad boost band as a function of the azimuth angle ϕ . The broad boost band is intended to be approximated by the magnitude of a filter, called “apprx”. This “apprx” filter is generated by a linear combination of the 2nd-order filters “low”, “mid”, “high” and “cut”. Hence the gains of the 2nd-order filters must be optimized, such that the error ϵ between the magnitude of the “apprx” filter $|H^{apprx}(f, \phi)|$ and the broad boost band is minimal. It is sufficient that this constraint is simultaneously respected at each of the center frequencies f_0^{low} , f_0^{mid} , f_0^{high} , and f_0^{cut} . Equation (4.5)

formulates this minimization problem, which has to be solved for each ϕ individually.

$$\min_{\hat{\mathbf{g}}} |\epsilon| = \min_{\hat{\mathbf{g}}} \left\| \begin{bmatrix} |H^{approx}(f_0^{low}, \phi)| \\ |H^{approx}(f_0^{mid}, \phi)| \\ |H^{approx}(f_0^{high}, \phi)| \\ |H^{approx}(f_0^{cut}, \phi)| \end{bmatrix} - \begin{bmatrix} g^{bbb}(\phi) \\ g^{bbb}(\phi) \\ g^{bbb}(\phi) \\ g^{cut}(\phi) \end{bmatrix} \right\| \quad (4.5)$$

$|H^{approx}(f_0^{low}, \phi)|$, $|H^{approx}(f_0^{mid}, \phi)|$, $|H^{approx}(f_0^{high}, \phi)|$, resp. $|H^{approx}(f_0^{cut}, \phi)|$ denotes the magnitude of the frequency response of the ‘‘apprx’’ filter at the frequencies $f = f_0^{low}$, $f = f_0^{mid}$, $f = f_0^{high}$, resp. $f = f_0^{cut}$ for the azimuth angle ϕ . $g^{bbb}(\phi)$ and $g^{cut}(\phi)$ define the target gains which $|H^{approx}(f, \phi)|$ is required to reach at the mentioned frequencies. The vector $\hat{\mathbf{g}}$ is of size 4×1 and its values comprise the gains for the 2nd-order filters such that the optimization problem is satisfied.

Abel and Berners proposed to use linear systems of equations to optimize the gains of succeeding peak filters such that the combination of the peak filters result in a transfer function with a magnitude of a given target gain [196]. For this we rewrite the optimization problem from Equation (4.5) as a linear system of equations, c.f. Equation (4.6), which is easily solved using least squares.

$$\mathbf{H} \cdot \hat{\mathbf{g}} = \mathbf{g} \quad (4.6)$$

The matrix \mathbf{H} is of size $n \times n$ where n equals the number of the involved peak filters, here $n = 4$. The elements $h_{i,j}$ ¹ of the matrix \mathbf{H} equal the magnitude of the frequency response $|H^j(f = f_0^i, \phi)|$ of the j^{th} peak filter for the azimuth angle ϕ , evaluated at the i^{th} center frequency, i.e. $h_{i,j} = |H^j(f = f_0^i, \phi)|$. The vectors \mathbf{g} and $\hat{\mathbf{g}}$ are of equal size (4×1). The elements g_j , resp. \hat{g}_i equal the magnitude of the target frequency response at the i^{th} center frequency, resp. the target gain of the j^{th} peak filter.

4.1.2.5 Filter combination

The individual steps of the parameterized spectral front-back difference HRTF, $\hat{\Delta}\text{HRTF}$, are illustrated in Figure 4.3. $\hat{\Delta}\text{HRTF}$ determines the center frequencies f_0 and bandwidth f_b , c.f. Equations (4.2) and (4.3) and the target gains g^{bbb} and g^{cut} , c.f. Equation (4.4). The optimized gains \hat{g} of the filters ‘‘low’’, ‘‘mid’’, ‘‘high’’, resp. ‘‘cut’’ are calculated accordingly to Equation (4.6). In order

¹When indexing matrix or vector elements (subscript): $i, j \in \{1, \dots, 4\}$. Otherwise (superscript): $j, i \in \{low, mid, high, cut\}$.

4.1. PARAMETRIC DIFFERENCE-HRTF

	0°/180°	22.5°/157.5°	45°/135°	67.5°/112.5°
meas	1.65 dB	1.41 dB	1.80 dB	0.99 dB
sim	1.31 dB	1.02 dB	1.29 dB	0.82 dB

Table 4.2: RMSE between the $\Delta HRTF$ based on the measurement (“meas”), resp. simulation (“sim”) and the $\hat{\Delta HRTF}$ in the frequency decade [1 kHz; 10 kHz].

to avoid discontinuities when the azimuth angle ϕ passes over the interaural axis, we introduced the output gain $\gamma_o(\phi)$ which is constantly +1, resp. -1 for $\phi \geq 80^\circ$, resp. $\phi \leq 100^\circ$. Between 80° and 100° , $g_o(\phi)$ follows a Tukey window symmetrical to $\phi = 90^\circ$, c.f. Figure 4.4 [76]. The gains \hat{g}^{low} , \hat{g}^{mid} , \hat{g}^{high} , \hat{g}^{cut} and g^{nar} are scaled with $\gamma_o(\phi)$. Finally, we obtained five 2nd-order IIR peak filters, defined by their center frequency f_0 , bandwidth f_b , and gain g_o . Their linear combination results in $\hat{\Delta HRTF}$ whose magnitude is shown in Figure 4.5.

The magnitudes of the measured, simulated, and parameterized spectral front-back difference HRTF are confronted in Figure 4.6 for the azimuth pairs of $0^\circ/180^\circ$ (see 0°), $22.5^\circ/157.5^\circ$ (see 22.5°), $45^\circ/135^\circ$ (see 45°), and $67.5^\circ/112.5^\circ$ (see 67.5°). Analysis of $\hat{\Delta HRTF}$ focused on the domain of definition ranging from 1 kHz to 10 kHz. By visual inspection it turns out that $\hat{\Delta HRTF}$ is closer to the smooth, simulated $\Delta HRTF$ than the fluctuating, measured $\Delta HRTF$. $\hat{\Delta HRTF}$ tends to overestimate the low and high frequency slope of the broad boost band, while it approximates well the amplitude of the broad boost band. The center frequency and the amplitude of the narrow peak are precisely reproduced by $\hat{\Delta HRTF}$. Table 4.2 summarizes the RMSE between $\hat{\Delta HRTF}$ and the measured, resp. simulated $\Delta HRTF$. The values show lower RMSE values for the simulated than for the measured $\Delta HRTF$, supporting the previously, visually determined similarity characteristics. As the pairs of azimuth angles become more lateral, the RMSE decreases. At $45^\circ/135^\circ$ this decrease shows discontinuities. They were caused by the large differences between the measured and simulated $\Delta HRTF$, particular in the center frequency of the narrow peak f_0^{nar} and in the high frequency slope of the broad boost band, c.f. Figure 4.6. The measured and simulated $\Delta HRTF$ show negative peaks near 1 kHz and 10 kHz which are not taken into account by $\hat{\Delta HRTF}$. The amplitudes of these negative peaks are converging towards zero with increasing lateral angles, leading to the general decreasing trend of the RMSE in Table 4.2.

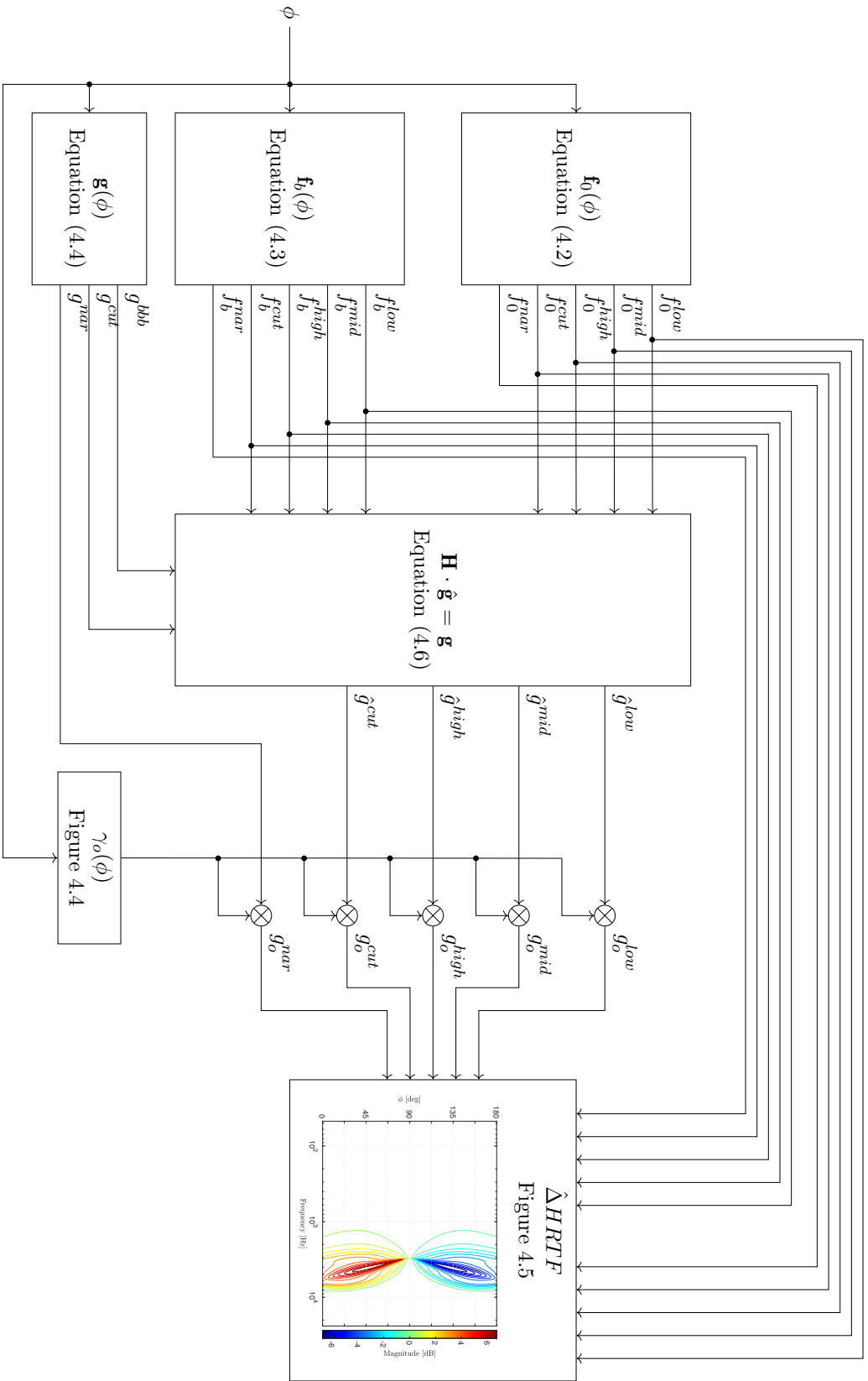


Figure 4.3: Structure of $\hat{\Delta}HRTF$. The input parameter ϕ controls the center frequencies f_0 , bandwidths f_b , and the gains g_o of the 2nd-order IIR filters “low”, “mid”, “high”, “cut”, and “nar”. The gains \hat{g} are determined by least square optimization. \hat{g} and the gain of the “nar” filter (g^{nar}) are scaled with the output gain $\gamma_o(\phi)$.

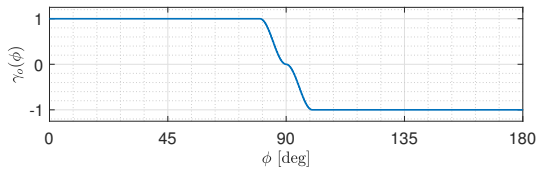


Figure 4.4: Output gain γ_o of $\hat{\Delta}$ HRTF filter as a function of the azimuth angle ϕ .

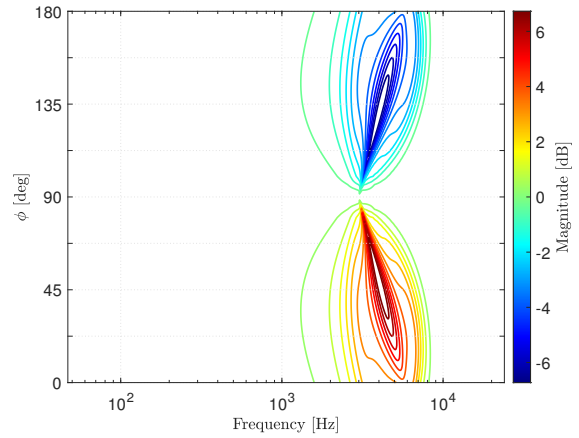


Figure 4.5: Contour lines of the magnitude of $\hat{\Delta}$ HRTF, as function of the frequency and azimuth angle ϕ .

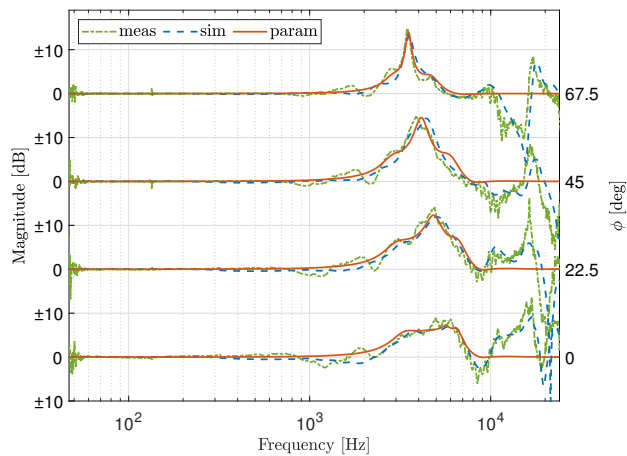


Figure 4.6: Comparison of $\hat{\Delta}$ HRTF (“param”) and the spectral front-back differences of the measured (“meas”) and simulated (“sim”) HRTFs of dummy head Harry33.

4.2 Listening Test

A listening test was conducted to assess in what extent listeners could interpret the provided spectral cues. For this, mono sounds were converted to stereo stimuli with directional information which were presented over headphones to the participants. Spectral cues were introduced to the stimuli by rendering filters. The localization performance obtained with the different rendering filters was evaluated. Two methods of generating the stimuli were used and it was evaluated how these methods influence the localization performance. Further, two sound samples were used to investigate if the spectral composition of the sound sample affects the localization performance.

4.2.1 Setup

Prior to the listening test, we performed a Békésy audiometry to verify that the listeners have normal hearing related to their age. The audiometry was conducted in the semi-anechoic chamber. We used pure tones, which ranged the interval [125 Hz; 8 kHz] in octave bands. The *Otometrics* Madsen Astera² clinical audiometer, TDH39 headphones, and the *Otometrics* Otosuite audiometry software were automatically controlling the test [179].

The listening test was conducted in an office pod (*SBS Silence Business Solutions* “Procyon Quattro”). Participants were seated and listened to stimuli via stereo headphones (*BeyerDynamics* “DT 770M”) which were driven by a tablet computer. The computer conducted the listening test and served as user interface. The interface resembled the user interface of the localization test, c.f. Figure 1.24c. Common elements were the acoustic horizon (black circle), the direction cursor (black ray), the progress bar and the button for validating the responses. The button was relocated outside the acoustic horizon and a head icon was placed at the center point of the acoustic horizon (ancient position of the response button), providing the reference orientation for frontal directions, c.f. Figure 4.7. The test was an unforced choice test, i.e. the direction cursor could be freely steered in any direction, whereas the step size was only limited by the technical specifications of the touch screen.

4.2.2 Stimuli

The stimuli were generated based on two mono sound samples. The first sample, i.e. male speech of 0.992 s, covered the lower frequency range up to approximately 3 kHz, c.f. Figure 4.8a. The second

4.2. LISTENING TEST

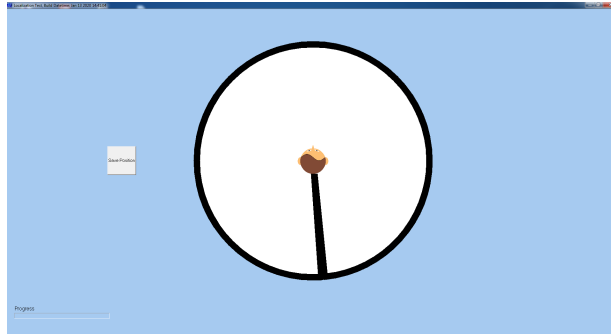


Figure 4.7: The GUI comprised of the progress bar, the direction cursor (black ray), the acoustic horizon (black circle), the reference orientation for the subject (head icon in the center), and the response validation button.

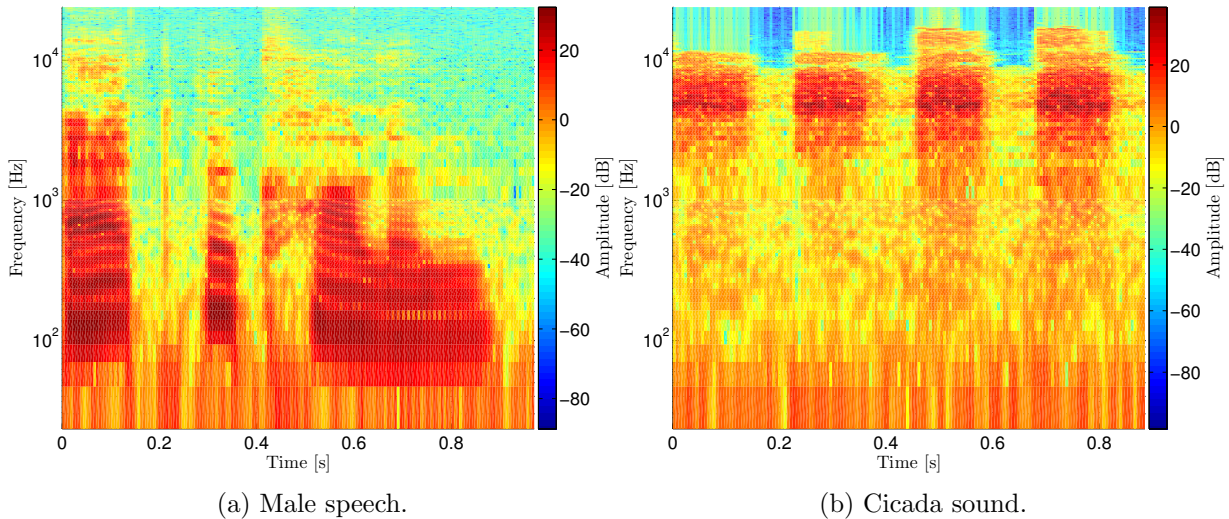


Figure 4.8: Spectrogram of the sound samples which were deployed during the listening test.

sample, i.e. cicada sound of 0.908 s, covered the high frequency range from approximately 2 kHz up to 10 kHz and resembled to pulsed noise, i.e. bursts, c.f. Figure 4.8a.

The methods of generating the stimuli were based on ambisonic rendering and discrete rendering and converted the mono sound sample into a stereo stimuli with directional information. The processing chains are sketched in Figures 4.9 and 4.10. Both methods included headphone equalization as terminating step.

The ambisonic method started by recording the sound field in the center point of the circular loudspeaker array with a 1st-order ambisonic microphone (TetraMic by *Core Sound*). The sound sample was positioned at discrete positions ϕ_t . The recorded A-format signals, i.e. output signals of

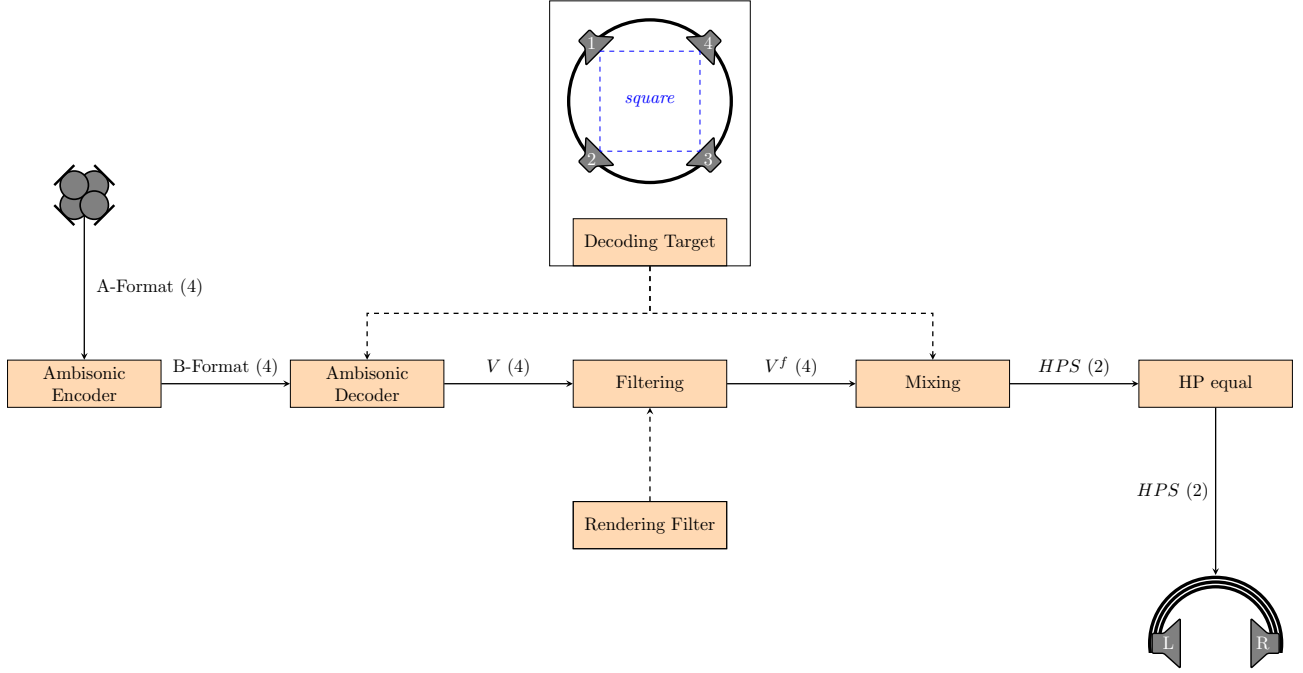


Figure 4.9: Processing chain for generating the stimuli using the ambisonic method. The signals are the virtual loudspeaker signal (V), the filtered, virtual loudspeaker signal (V^f), and the headphone signal (HPS). The digits in parenthesis denote the number of channels per signal.

the microphone capsules, were encoding to B-format signals, called W, X, Y, and Z. The B-format signals were decoded for a squared virtual loudspeaker setup, c.f. “Decoding Target” in Figure 4.9, resulting in the virtual loudspeaker signals V_1 to V_4 . Spectral front-back information was introduced to the virtual loudspeaker signals by applying rendering filters, c.f. “Rendering Filter” in Figure 4.9. The filtered, virtual loudspeaker signals V_1^f to V_4^f were down mixed to a stereo headphone signal HPS , accordingly to Equation (4.7), and finally headphone equalization was performed.

$$HPS_L = V_1^f + V_2^f \quad HPS_R = V_3^f + V_4^f \quad (4.7)$$

Rendering the stimuli by the discrete method, c.f. Figure 4.10, started by converting the mono sound sample into a stereo sound sample and introducing the spectral front-back information by applying rendering filters, c.f. “Rendering Filter” in Figure 4.10. The values of $r = 0.105$ m (head diameter) and $c = 343.23$ m s^{-1} (speed of sound) were inserted into the approximation of the ITD as given in Equation (1.3). The resulting expression, c.f. Equation (4.8), defined the ITD and the left and right channels were correspondingly delayed to each other. Finally, headphone equalization was

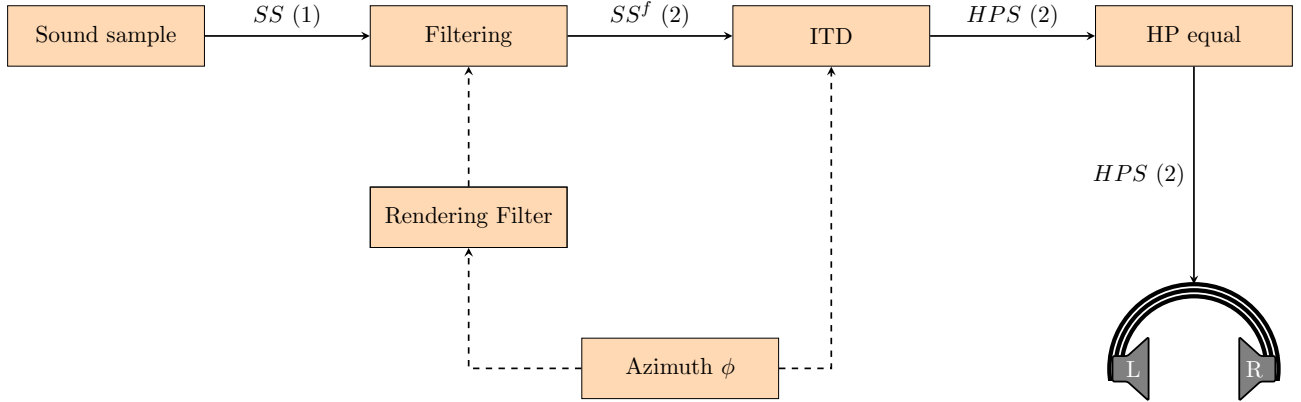


Figure 4.10: Processing chain for generating the stimuli using the discrete method. The signals are the sound sample (SS), the filtered sound sample (SS^f), and the headphone signal (HPS). The digits in parenthesis denote the number of channels per signal.

performed.

$$ITD(\phi_t) = 611.83 \mu s \cdot \sin(\phi_t). \quad (4.8)$$

4.2.2.1 Rendering Filters

If not explicitly otherwise stated, the rendering filters differentiated only between frontal and back sound incidence but didn't differentiate between the left and right ear. Hence, their frequency responses were symmetric to the median plane, i.e. $H(f, \phi) = H(f, 360^\circ - \phi)$. In the ambisonic rendering method, the virtual loudspeaker signals V_1 and V_4 , resp. V_2 and V_3 are filtered with the “frontal”, resp. “back” frequency response. In the discrete rendering method, the sound sample was filtered with the “frontal”, resp. “back” frequency response for $\phi_t < 90^\circ$ or $\phi_t > 270^\circ$, resp. $90^\circ < \phi_t < 270^\circ$.

4.2.2.1.1 High shelf filter According to Frank et al., a 3kHz HSF enhances front-back discrimination [76]. They reported that a filter gain of ± 6 dB leads to an unnaturally sharp sound but good localization results. With the implementation of 2nd-order HSFs from [194] we obtained the frequency response for frontal and back sound incidence as shown in Figure 4.11. In the following, filter N° 1 refers to this filter.

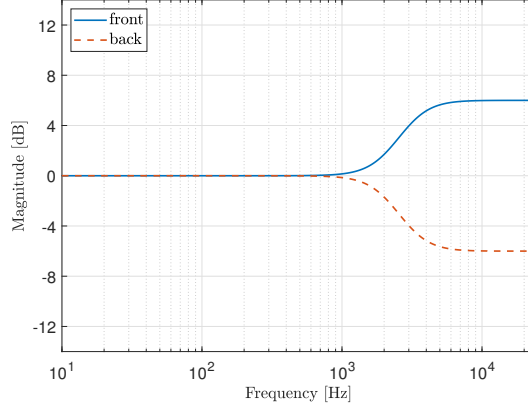


Figure 4.11: 3 kHz HSF (filter N° 1).

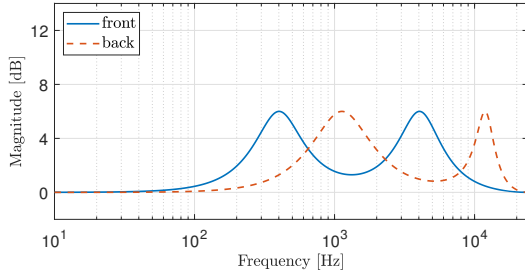
4.2.2.1.2 Blauert band filters Each coherent frequency interval of Blauert’s directional and boosted bands of 90 % and 95 % confidence, as given in Table 1.1, was represented by one 2nd-order peak filter. The center frequencies f_0 , resp. bandwidths f_b of the peak filters were calculated accordingly to Equation (4.9), resp. Equation (4.10). f_l , resp. f_u denote the lower, resp. upper frequency of a coherent frequency interval. The IIR implementation from [194] was used. Peak filters which correspond to the same type of band were linearly combined with the optimization technique from [196], presented in Section 4.1.2.4. The optimization was performed with a target gain of 6 dB. Figure 4.12 visualizes the magnitudes of the obtained filters, which are referred to as filter N° 2 to N° 5.

$$f_0 = \sqrt{f_l^2 \cdot f_u^2} \quad (4.9)$$

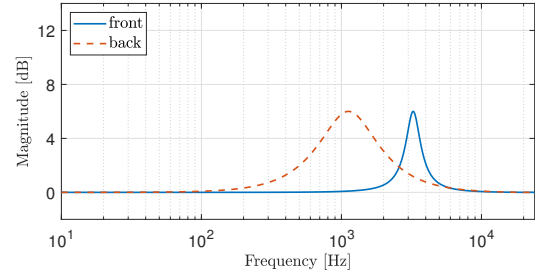
$$f_b = f_u - f_l \quad (4.10)$$

4.2.2.1.3 Parameterized difference-HRTF $\hat{\Delta}$ HRTF, i.e. the parameterized spectral front-back difference HRTF as developed in Section 4.1.2, was evaluated for the ambisonic render at $\phi \in \{45^\circ, 135^\circ\}$ and for the discrete method at $\phi = \phi_t$. In the ambisonic method, loudspeaker signals V_1 and V_4 , resp. V_2 and V_3 were filtered with the “frontal” ($\phi = 45^\circ$), resp. “back” ($\phi = 135^\circ$) $\hat{\Delta}$ HRTF, as visualized in Figure 4.13a. In the discrete method $\hat{\Delta}$ HRTF was evaluated for all possible azimuth positions ϕ_t . Pairs of azimuth angles symmetrical to the median plane, led to equal filters, see Figure 4.13b. Filter N° 6 refers to the $\hat{\Delta}$ HRTF based filter.

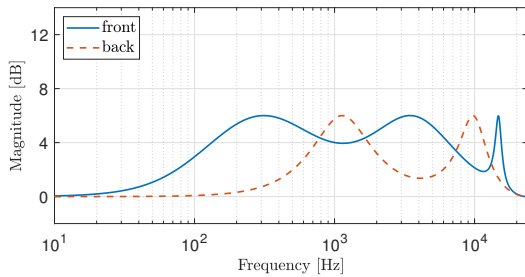
4.2. LISTENING TEST



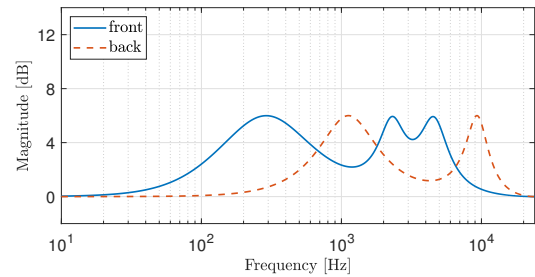
(a) Directional band 90 % (filter N° 2).



(b) Directional band 95 % (filter N° 3).

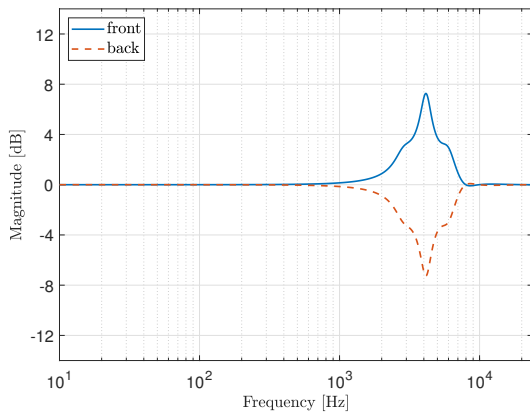


(c) Boosted band 90 % (filter N° 4).

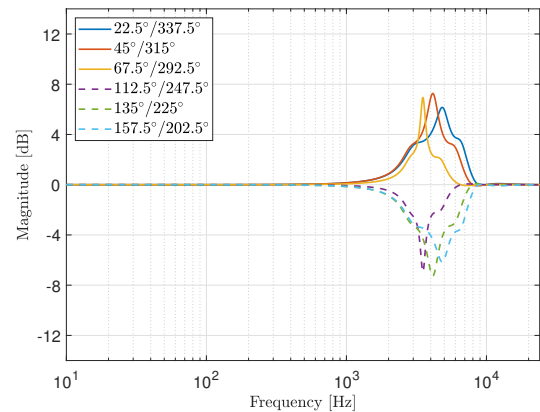


(d) Boosted band 95 % (filter N° 5).

Figure 4.12: Filters, which are based on Blauert's directional and boosted bands of 90 % and 95 % confidence, were implemented with 2nd-order peak filters.



(a)



(b)

Figure 4.13: $\hat{\Delta}$ HRTF (filter N° 6) as applied to the signals of front and back virtual loudspeakers in the ambisonic rendering method (Figure 4.13a) and as applied to the sound sample in the discrete rendering method (Figure 4.13b).

4.2. LISTENING TEST

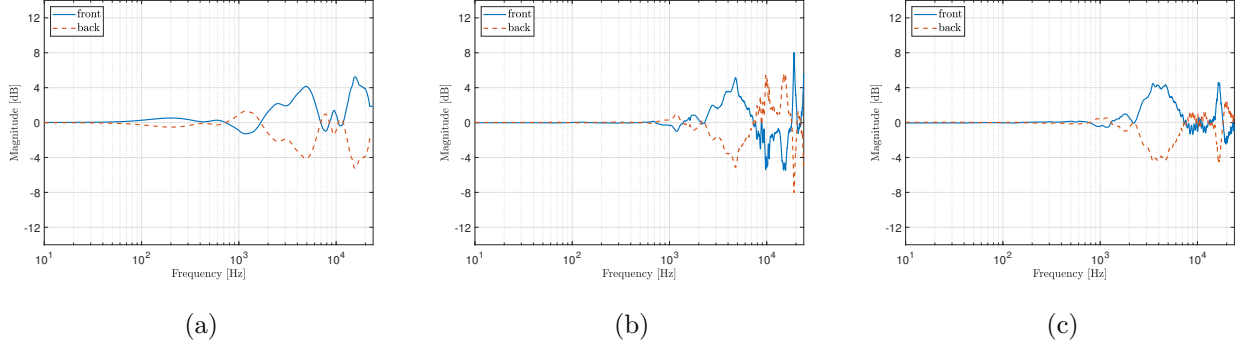


Figure 4.14: $\overline{\Delta\text{HRTF}}$ as based on the average HRTF of the CIPIC database (filter N° 7, Figure 4.14a), the measured HRTFs of dummy head Harry34 (filter N° 8, Figure 4.14b), and the measured HRTFs of dummy head Harry33 (filter N° 9, Figure 4.14c).

4.2.2.1.4 Averaged difference-HRTF We defined the averaged, spectral front-back difference HRTF, called $\overline{\Delta\text{HRTF}}$. Based on the magnitude of the initial HRTF, called $\text{HRTF}(f, \phi)$, we calculated the average HRTF, called $\overline{\text{HRTF}}(f, Q_i)$, over the frontal ipsilateral quadrant ($i = 1$) and the back ipsilateral quadrant $i = 2$, c.f. Equation (4.11). The front-back difference was obtained by calculating the difference between both quadrants and scaling the result with 0.5, c.f. Equation (4.12).

$$\overline{\text{HRTF}}(f, Q_i) = \frac{1}{3} \cdot \sum_{\phi \in \Phi_i} |\text{HRTF}(f, \phi)| \quad \forall i \in \{1, 2\} \quad (4.11)$$

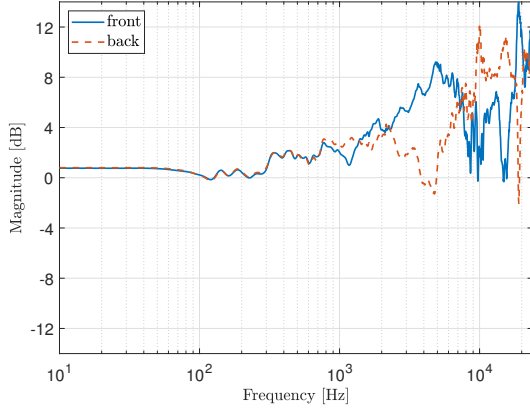
$$\text{with } \Phi_i = \begin{cases} \{22.5^\circ, 45^\circ, 67.5^\circ\}, & \text{for } i = 1 \\ \{112.5^\circ, 135^\circ, 157.5^\circ\}, & \text{for } i = 2 \end{cases}$$

$$\overline{\Delta\text{HRTF}}(f) = \frac{1}{2} \cdot (\overline{\text{HRTF}}(f, Q_1) - \overline{\text{HRTF}}(f, Q_2)) \quad (4.12)$$

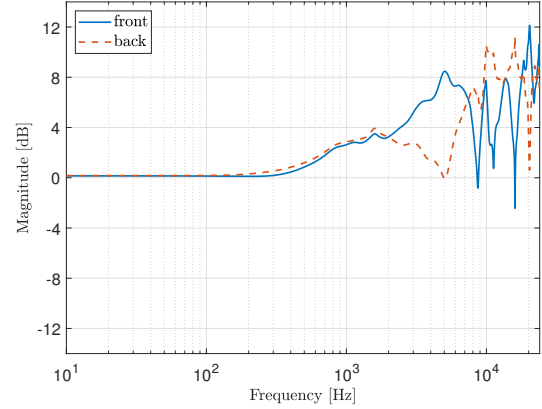
As initial HRTF we took the average HRTF of the CIPIC database [61] (resulting in filter N° 7) and the measured HRTFs of dummy head Harry34 (resulting in filter N° 8) and Harry33 (resulting in filter N° 9). The obtained frequency responses for frontal and back sound incidence are visualized in Figure 4.14.

4.2.2.1.5 Averaged HRTF From the calculation of the $\overline{\Delta\text{HRTF}}$ in Paragraph 4.2.2.1.4, we reused the definition of the quadrant wise averaged HRTF, i.e. $\overline{\text{HRTF}}(f, Q_i)$. HRTF data which were related to the frontal ipsilateral quadrant Q_1 , resp. the back ipsilateral quadrant Q_2 were defining the rendering

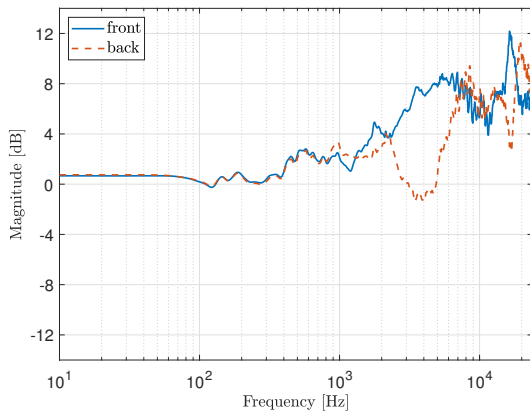
4.2. LISTENING TEST



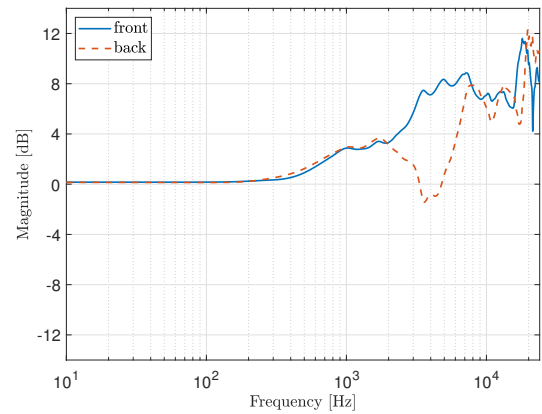
(a) Harry34, measured (filter N° 10).



(b) Harry34, simulated (filter N° 11).



(c) Harry33, measured (filter N° 12).



(d) Harry33, simulated (filter N° 13).

Figure 4.15: $\overline{\text{HRTF}}$ of the measured (Figures 4.15a and 4.15c) and simulated (Figures 4.15b and 4.15d) HRTF of dummy head Harry34 and Harry33.

filters for frontal, resp. back virtual loudspeaker signals or azimuth positions ϕ . $\overline{\text{HRTF}}$ was calculated for the measured and simulated HRTFs of the dummy heads Harry34 and Harry33. The resulting filters are referred to as filter N° 10 to N° 13. The obtained frequency responses for frontal and back positions are shown in Figure 4.15.

4.2.2.1.6 Simulated HRTF The rendering filter N° 14 was based on the simulated HRTF of dummy head Harry33, c.f. Section 3.1. With this rendering filter, the signal processing changed slightly as it took into account the ipsilateral and contralateral data of the HRTF. In the ambisonic rendering method, the four virtual loudspeaker signals V_1 to V_4 were binaurally filtered with the HRTF, c.f.

4.2. LISTENING TEST

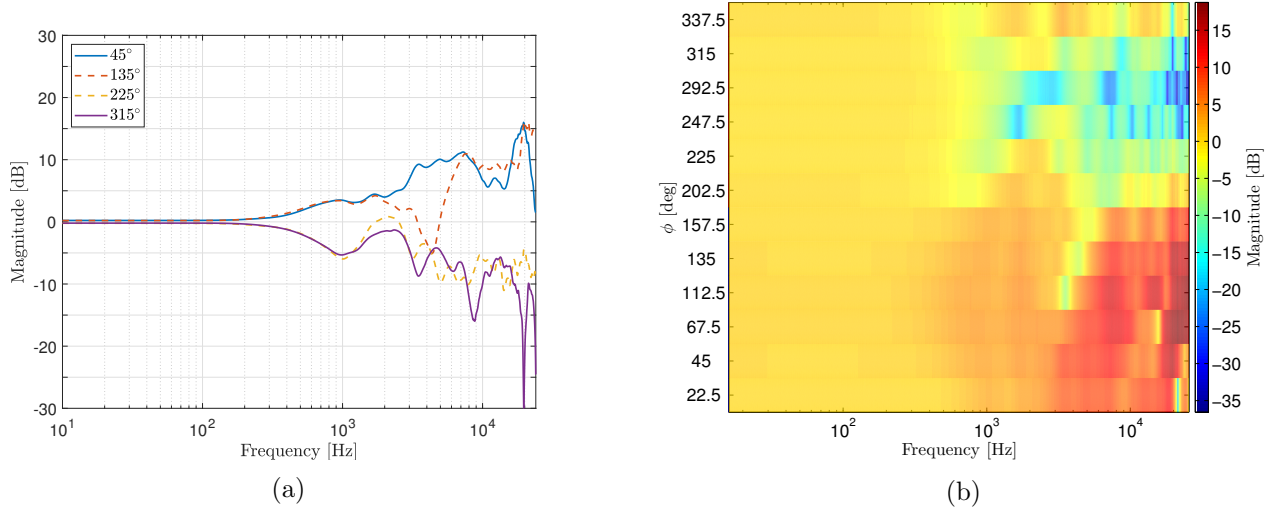


Figure 4.16: HRTF based rendering filter for the ambisonic (Figure 4.16a) and discrete (Figure 4.16b) method.

Equation (4.13). The frequency responses of the simulated HRTF aligning with the the positions of the virtual loudspeakers are shown in Figure 4.16a. In the case of the discrete rendering, the stereo sound sample was binaurally filtered with the simulated HRTF. The simulated HRTF is shown in Figure 4.16b for an exemplary set of azimuth angles ϕ .

$$\begin{bmatrix} HPS_L \\ HPS_R \end{bmatrix} = \begin{bmatrix} V_1 & V_2 & V_3 & V_4 \\ V_4 & V_3 & V_2 & V_1 \end{bmatrix} \cdot \text{HRTF} \left(f, \begin{bmatrix} 45^\circ \\ 135^\circ \\ 225^\circ \\ 315^\circ \end{bmatrix} \right) \quad (4.13)$$

4.2.3 Procedure

The stimuli were presented in complete random order to the listeners and each stimulus was tested twice. To keep the test duration at an acceptable length, i.e. below 20 min, the test was divided into two series. Subjects were allowed to participate in one of the series or in both series.

In the 1st series we tested stimuli based on both audio samples generated by the ambisonic method in combination with the rendering filters N° 1 to N° 13. Due to the large number of filters only few sound directions were tested. Stimuli were located at $\phi_t = \{0^\circ, 67.5^\circ, 180^\circ, 247.5^\circ\}$, covering the horizontal plane approximately uniformly. The directions of 67.5° and 247.5° replaced the directions of 90° and 270°, which were intended to be initially tested. The cones of confusions of 67.5° and 247.5°

4.2. LISTENING TEST

Filter N°	Description	Method	Series
1	3 kHz HSF	amb	1
		dis	2
2	Blauert’s directional band, 90 % confidence	amb	1
3	Blauert’s directional band, 95 % confidence	amb	1
4	Blauert’s boosted band, 90 % confidence	amb	1
5	Blauert’s boosted band, 95 % confidence	amb	1
		dis	2
6	Parameterized difference HRTF	amb	1
		dis	2
7	Averaged difference HRTF, KEMAR	amb	1
8	Averaged difference HRTF, Harry34, measured	amb	1
9	Averaged difference HRTF, Harry33, measured	amb	1
10	Averaged HRTF, Harry34, measured	amb	1
11	Averaged HRTF, Harry34, simulated	amb	1
12	Averaged HRTF, Harry33, measured	amb	1
13	Averaged HRTF, Harry33, simulated	amb	1
14d	HRTF, Harry33, simulated	dis	2
14s	HRTF, Harry33, simulated	amb	2

Table 4.3: Summary of the rendering filters. The separations correspond to the paragraphs in Section 4.2.2.1. The processing methods are ambisonic (“amb”) and discrete (“dis”).

are three dimensional, whereas the cones of confusions of 90° and 270° collapse to a plane. Subjects of the 1st series listen in total to 208 stimuli (13 filters, 4 directions, 2 sound samples, and 2 repetitions).

In the 2nd series we augmented the number of sound directions. We tested stimuli based on the male speech only. Stimuli were generated by the ambisonic method in combination with the rendering filter N° 14 and by the discrete method in combination with the rendering filters N° 1, N° 5, N° 6, and N° 14. The stimuli were located at $\phi_t = \{22.5^\circ, 45^\circ, \dots, 337.5^\circ\} \setminus \{90^\circ, 180^\circ, 270^\circ\}$, covering nearly the entire horizontal plane. Directions of 90° and 270° were excluded as they lack front-back pairs, while 0° and 180° were already included in the 1st series. Subjects listened in total to 120 stimuli (5 filters, 12 directions, 1 sound sample, and 2 repetitions).

None of the series provided an adaptation nor training phase. After an introduction to the test by the supervisor, subjects started directly with the listening test. Table 4.3 summarizes the deployed combinations of rendering filters and processing methods for both series.

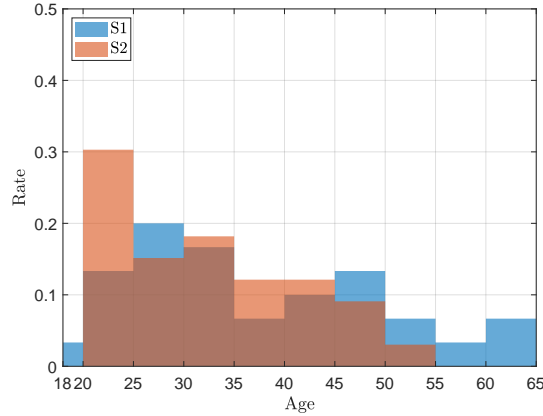


Figure 4.17: Distribution of the age of the participants of the 1st series (S1) and the 2nd series (S2).

4.2.4 Participants

In total, 45 subjects with normal hearing participated in the test. The 1st and 2nd series were performed by 30 and 33 subjects, and 18 subjects performed both series. The females-to-males ratio in both groups was 1 : 2. The distribution of the subjects' age is shown in Figure 4.17. Subjects of the 1st, resp. 2nd series were between 19 and 64 (mean: 36.5 years, median: 34 years), resp. 20 and 54 years old (mean: 32.09 years, median: 31 years). The population of participants was noticeably younger in the 2nd series (30 % below age 25) than in the 1st series (16.67 % below age 25). In contrast, the age of the subjects of the 1st series ranged a larger interval than the age of the subjects of the 2nd series.

4.2.5 Results

A preliminary evaluation has shown that the filters of the 1st series show very similar results. To facilitate the evaluation of the 1st series, a representative filter was used for each paragraph of Section 4.2.2.1. The filters N° 1, 5, 6, 9, and 12 were selected. These filters were either part of the 2nd series or were based on the same dummy head as filters of the 2nd series. Evaluation of the 2nd series was done on all deployed filters.

During the 2nd series the ambisonic and discrete rendering methods were used in combination with filter N° 14. To present the data in a more compact way, we defined the filters N° 14s and N° 14d. Filter N° 14s refers to the ambisonic rendering method (“s”: squared decoding target) and filter N°

14d refers to the discrete rendering method (“d”: discrete rendering).

4.2.5.1 Response time

The average response time was 2.37 s, resp. 2.40 s during the 1st, resp. 2nd series. Figures 4.18a and 4.18b show the distribution of the response time across the rendering filters. The response time t_r varied between 2.33 s and 2.42 s across the rendering filters of the 1st series. No significant effect of the rendering filter on the response time was identified by performing a Kruskal-Wallis test (p-value $\gg 0.05$). In the 2nd series, t_r varied across the rendering filters between 2.22 s and 2.90 s. Contrarily to the 1st series, a Kruskal-Wallis test proved that the rendering filter had a significant effect on the response time during the 2nd series (p-value: $\ll 0.001$). With a pairwise Kruskal-Wallis test we identified that rendering filters N° 14s led to significantly increased t_r than rendering filters N° 1, N° 5, N° 6, and N° 14d. Between rendering filters N° 1, N° 5, N° 6, and N° 14d no difference was identified.

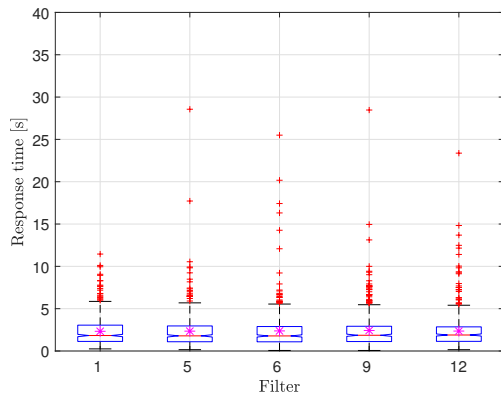
The trend of t_r along the test progress and its moving average showed equal characteristics in both series, c.f. Figures 4.18c and 4.18d. After familiarizing with the test during the first 3% of the stimuli, subjects speed up with responding. Their response time converged to 2 s above 30% of the test progress.

4.2.5.2 General performance

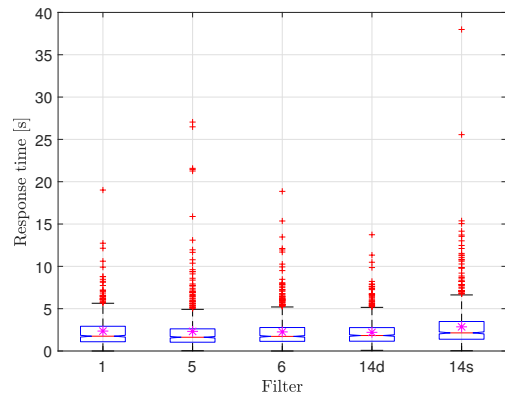
Figure 4.19 visualizes the confusion matrices for both series. Perfect localization is highlighted by the main diagonal and front-back confusions are highlighted by anti-diagonals. The discrete tested angles ϕ_t are plotted on the horizontal axis and the responded angles ϕ_r are plotted on the vertical axis. For increased visibility we discretized the continuous scale of ϕ_r (remember that the data are obtained with an unforced choice test) with a step size of 11.25° . It is obtained that for both series only little data is located on the main diagonal.

In the 1st series, the responded angles ϕ_r are spread over the entire horizontal plane independent of ϕ_t , see Figure 4.19a. Stimuli which are positioned on the median plane ($\phi_t \in \{0^\circ, 180^\circ\}$) were localized in 37.25% at frontal median positions ($\phi_r = 0^\circ$) but only in 20.67% at back median positions ($\phi_r = 180^\circ$). Stimuli which were located at $\phi_t = 67.5^\circ$ and $\phi_t = 247.5^\circ$ are more frequently localized at $\phi_r = 90^\circ$ (35.06%) and $\phi_r = 270^\circ$ (37.63%) than at the correct positions of $\phi_r = 67.5^\circ$ (7.12%) and $\phi_r = 247.5^\circ$ (12.05%).

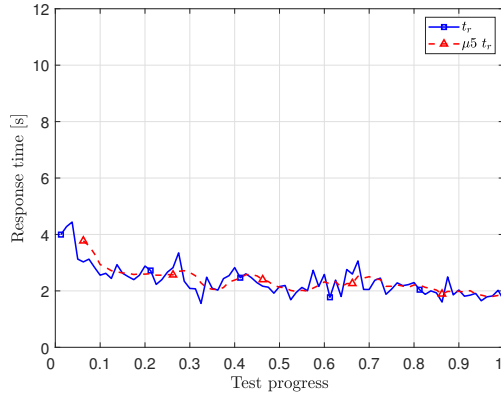
4.2. LISTENING TEST



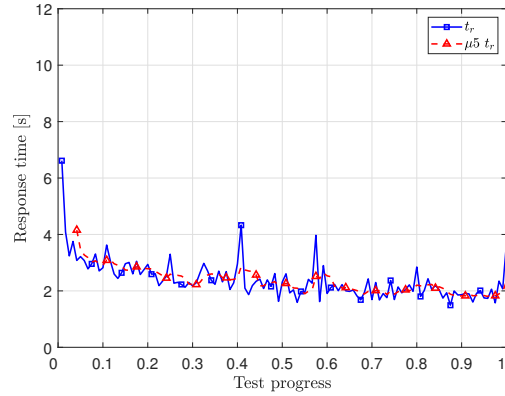
(a)



(b)



(c)



(d)

Figure 4.18: Box plot with mean values (asterisks) of the response times across the rendering filters during the 1st series (Figure 4.18a) and 2nd series (Figure 4.18b). The trend of the averaged response time (t_r) and its moving average over 5 samples ($\mu_5 t_r$) along the test progress is shown in Figure 4.18c (1st series) and Figure 4.18d (2nd series).

4.2. LISTENING TEST

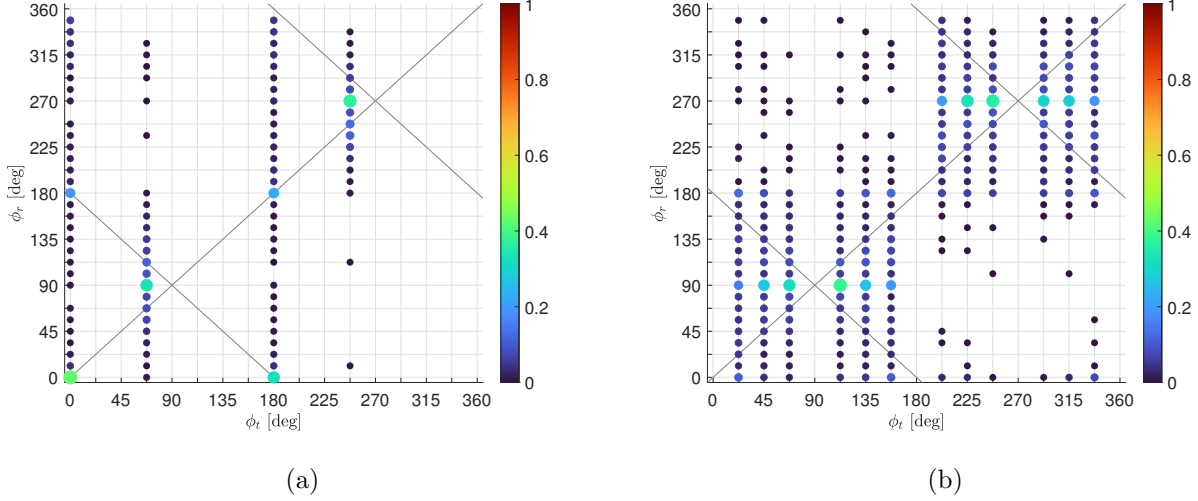


Figure 4.19: Confusion matrices of the 1st series (Figure 4.19a) and 2nd series (Figure 4.19b). The distributions of the responded angels ϕ_r are visualized over the tested angles ϕ_t . The size and the color of the circles indicate the relative frequency of the pairs of the test angle ϕ_t and responded angle ϕ_r . The latter was discrete with a step size of 11.25° .

Regarding the confusion matrix of the 2nd series in Figure 4.19b, the data is mainly located in the lower left or upper right submatrix with some outliers in the lower right and upper left submatrix. Further, stimuli were often localized at lateral positions independently of ϕ_t . We notice the trend that the more lateral, resp. median a stimulus is positioned, the more frequent subjects localized the stimulus at $\phi_r \in \{90^\circ, 270^\circ\}$, resp. $\phi_r \in \{0^\circ, 180^\circ\}$. The distributions of ϕ_r show nearly constant characteristics in the left and right hemisphere with peaks at lateral and median positions.

4.2.5.3 Angular error

The distributions of the signed angular error ϵ of the 1st and 2nd series are visualized in Figure 4.20a with a resolution of 5.625° . Both distributions show approximately the characteristic of even functions, i.e. they are symmetric with respect to $\epsilon = 0^\circ$. Focusing on the 1st series, the peak of the distribution is located at $\epsilon = 0^\circ$ with an amplitude of 27.88%. Two minor peaks are located at $\epsilon = \pm 22.5^\circ$ with an amplitude of 7.72% and 9.60%. Outside the interval of $\epsilon = \pm 22.5^\circ$ the distribution drops to 1.12% and 2.1% and decreases towards zero. Regarding the 2nd series, the distribution shows a repetitive pattern with local maxima at integer multiples of $\epsilon = \pm 22.5^\circ$ and local minima at odd multiples of $\pm 11.25^\circ$. The rate of signed angular errors outside $\epsilon = \pm 67.5^\circ$ converges towards zero.

4.2. LISTENING TEST

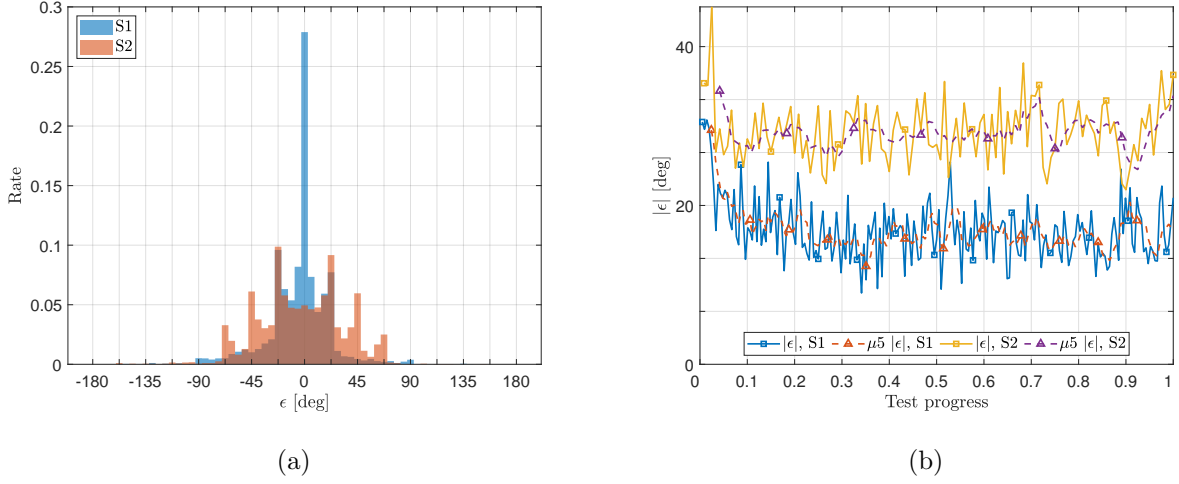


Figure 4.20: Distribution of the signed angular error ϵ (Figure 4.20a), the trend of the unsigned angular error $|\epsilon|$ along the test progress and its moving average over 5 samples $\mu_5 |\epsilon|$ (Figure 4.20b) of the 1st series (S1) and the 2nd series (S2).

The trend of the unsigned angular error $|\epsilon|$, averaged over all subjects, and its moving average over 5 samples are traced for both series in Figure 4.20b. It is noticed that the curves are shifted by approximately 13.33° . $|\epsilon|$ varies during the 1st, resp. 2nd series around its mean value of 16.88° , resp. 29.36° with a standard deviation of 3.80° , resp. 3.77° . After initially high values for $|\epsilon|$ the curves converge within the first 3% of the stimuli and vary constantly around their mean values.

Regarding the two sound samples of the 1st series, we obtained an unsigned angular error of 16.11° (male speech) and 17.65° (cicada sound). This difference was statistically significant, accordingly to a Kruskal-Wallis test (p-value: < 0.01).

The distributions of the unsigned angular error $|\epsilon|$ with respect to the tested angles ϕ_t are visualized in Figure 4.21 with box plots. For both series we notice that the range of the outliers is smaller at median positions than at lateral positions. In the 1st series, the average of $|\epsilon|$ varied little with ϕ_t but the median values are highly dependent on ϕ_t . In the 2nd series, the average values of $|\epsilon|$ per direction increased for median positions and decrease for lateral positions.

4.2.5.4 Correct responses

With the definition of the correct responses, c.f. Section 2.3.3.3, and the tight condition, i.e. $\phi_r \in \left[\phi_t - \frac{22.5^\circ}{2}; \phi_t + \frac{22.5^\circ}{2} \right]$, we obtained a correct response rate of 27.29% (1st series) and 9.89%

4.2. LISTENING TEST

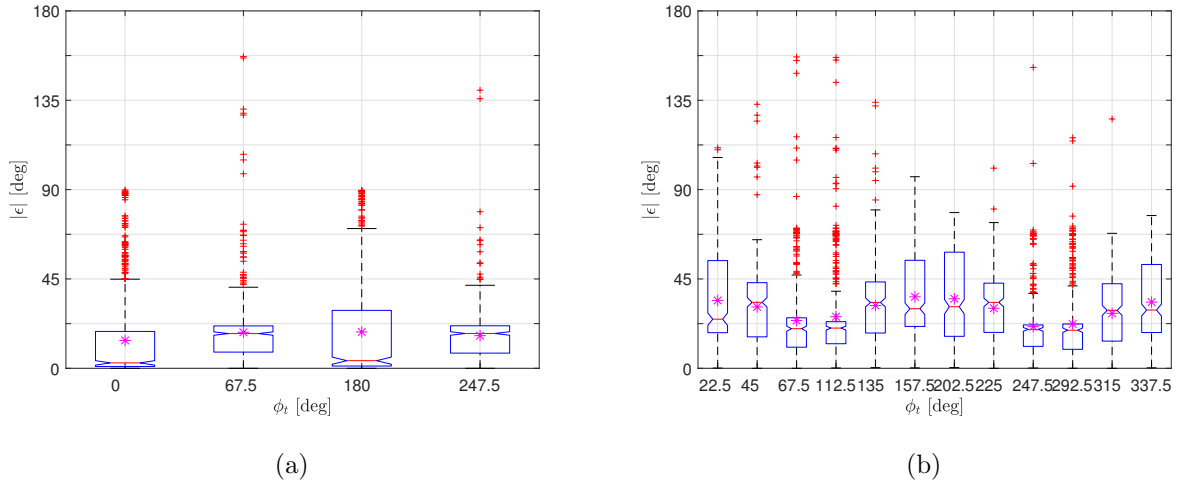


Figure 4.21: Box plots with mean values (asterisks) of the unsigned angular error $|\epsilon|$ for the tested angles ϕ_t , as obtained during the 1st series (Figure 4.21a) and the 2nd series (Figure 4.21b).

(2nd series). The correct response rate of the 1st series depended significantly on the sound samples (Standard ANOVA resulted in p -value < 0.05) and varied between 39.75 % (male speech) and 25.83 % (cicada sound).

For filters which are included in the 1st series, resp. 2nd series the correct response rate varies in the interval [24.37 %; 30.63 %], resp. [7.08 %; 12.64 %]. Figures 4.22a and 4.22b break down the correct response rate for each rendering filter. By visual inspection of the box plots, it is obtained that almost all notches of the boxes overlap, which means that the filters had no statistically significant effect on the correct answer rate. This was supported by a standard one-way ANOVA (p -values: > 0.05).

Regarding the discrete processing method in the 2nd series most, second most, and fewest correct responses were obtained with rendering filters N° 14d (12.64 %), N° 1 (10.97 %), and N° 6 (9.03 %), respectively. The simulated, non-individual HRTF, i.e. filter N° 14d, led to at least 1.67 % more correct responses than boosting selected frequency intervals, i.e. filters N° 1, N° 5, and N° 6. The different processing methods, which were deployed in the 2nd series with the filters N° 14d and N° 14s, led to correct response rates of 12.64 % (N° 14d) and 7.08 % (N° 14s). The notches in the interquartile range are not overlapping and a standard ANOVA proved that the obtained difference is statistically significant (p -value: < 0.05).

The trend of the correct response rate along the test progress and its moving average over 5 samples are drawn in Figures 4.22c and 4.22d. The graphs of the 1st and 2nd series do not show any noticeable

trend. An interval of familiarization, as it was obtained with the response time (Section 4.2.5.1) and the absolute angular error (Section 4.2.5.3), is not obtained with the correct response rate.

4.2.5.5 Confusion rate

Front-back, left-right and mixed confusions were calculated based on the definitions given in Section 2.3.3.5. Hence, stimuli where ϕ_t aligned with the frontal plane, median plane, resp. frontal or median plane weren't considered. As a result, the left-right and mixed confusion rates of the 1st series didn't considered stimuli with $\phi_t \in \{0^\circ, 180^\circ\}$, i.e. they are calculated on only 50% of the stimuli that were presented during the 1st series. The 2nd series was not affected because it didn't contain stimuli at $\phi_t \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$.

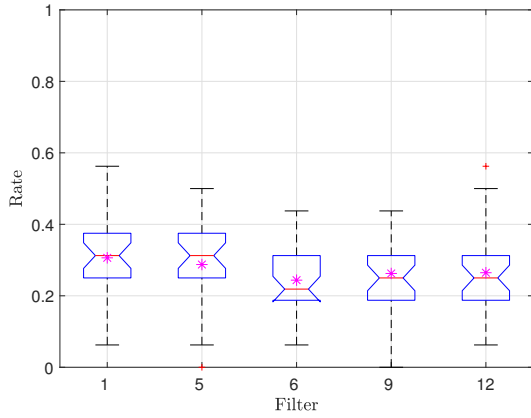
The distributions of the front-back, left-right and mixed confusions for the rendering filters of the 1st, resp. 2nd series are given in Figures 4.23a to 4.23c, resp. Figures 4.23d to 4.23f. The results are presented in detail in the following.

4.2.5.5.1 Front-back confusion The front-back confusion rate was in average 31.83% in the 1st series and 43.78% in the 2nd series.

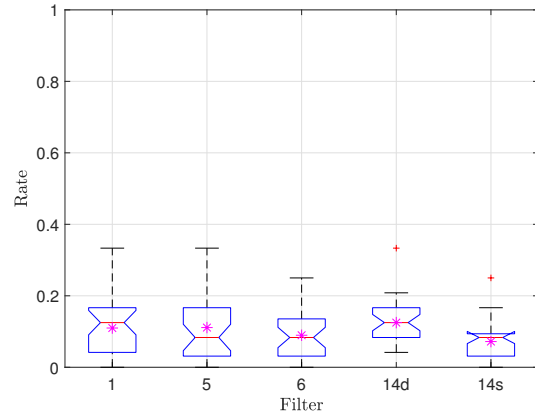
The average front-back confusion rate for filters of the 1st series span a narrow interval between [30.21%; 34.38%], c.f. asterisks in Figure 4.23a. According to the overlapping notches and a standard ANOVA, the differences in the averaged front-back confusion rates between the filters was not significant (p-value: $\gg 0.05$). In the 2nd series, the average front-back confusion rates of the individual filters span the interval [36.25%; 52.22%]. A standard ANOVA showed that the filters had a significant effect on the number of front-back confusions (p-value: $\ll 0.001$). A pairwise ANOVA revealed significant differences between filter N° 14d and the other four filters (pairwise p-values: < 0.001) and between filter N° 6 and N° 14s (p-value < 0.01). Grouping the filters accordingly to the 0.001 level of significance, results in the groups of {N° 1, N° 5, N° 6, N° 14s} and {N° 14d}.

In both series the tested azimuth angle ϕ_t showed significant effects on the number of front-back confusions (p-values: $\ll 0.001$). In the 1st series we noticed increased front-back confusion rates for lateral positions of $\phi_t \in \{67.5^\circ, 247.5^\circ\}$ (confusion rate $\geq 33.83\%$) than for median positions of $\phi_t \in \{0^\circ, 180^\circ\}$ (confusion rate $\leq 24.17\%$). In the 2nd series frontal stimuli ($\phi_t \leq 67.5^\circ$, or $\phi_t \geq 247.5^\circ$) led to front-back confusions, e.g. front-to-back reversals, in at least 43.33% while back stimuli ($\phi_t \in$

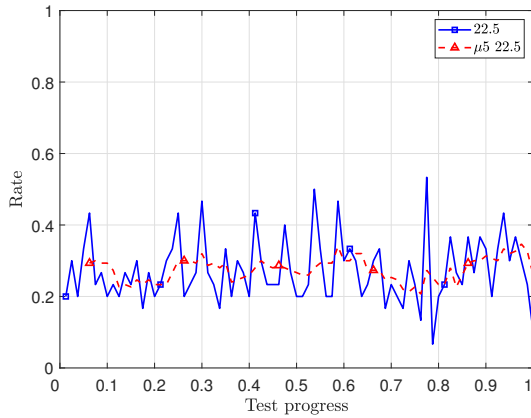
4.2. LISTENING TEST



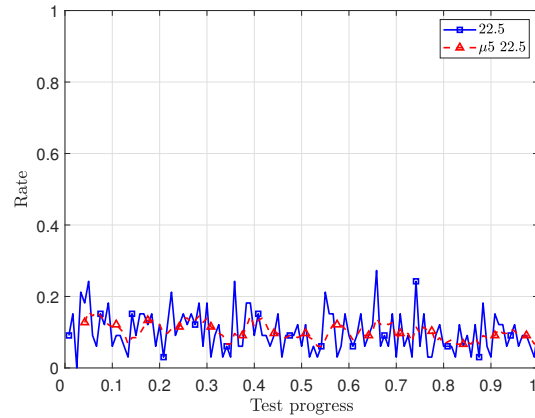
(a) 1st series.



(b) 2nd series.



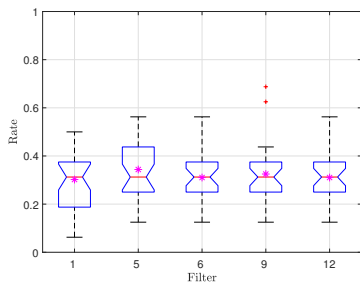
(c) 1st series.



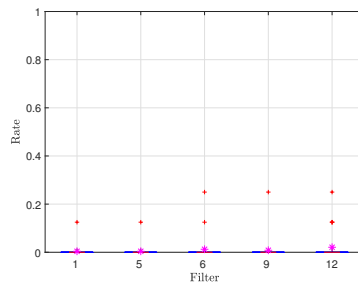
(d) 2nd series.

Figure 4.22: Box plots with mean values (asterisks) of the correct response rate under the tight condition ($\phi_r \geq \phi_t \mp \frac{22.5^\circ}{2}$) are shown per filter (Figures 4.22a and 4.22b). The trend of the correct response rate and its moving average over 5 samples (μ_5) along the test progress is shown in (Figures 4.22c and 4.22d).

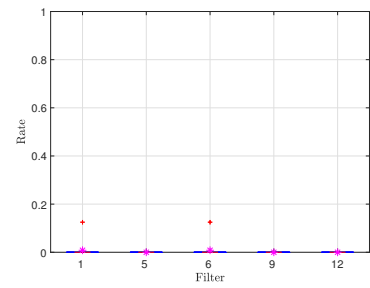
4.2. LISTENING TEST



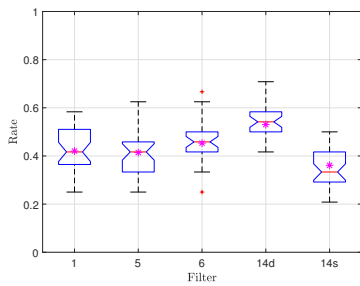
(a) Front-back confusions.



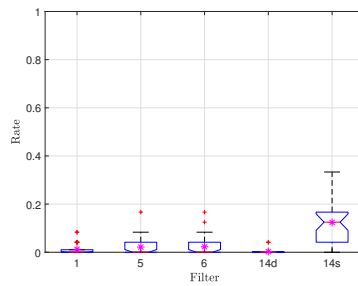
(b) Left-right confusions.



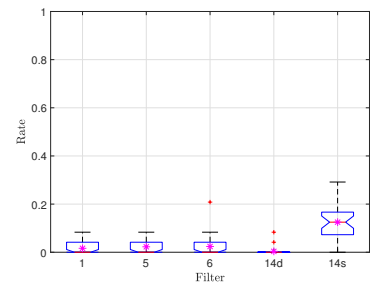
(c) Mixed confusions.



(d) Front-back confusions.



(e) Left-right confusions.



(f) Mixed confusions.

Figure 4.23: Box plots with mean values (asterisks) of the confusion rates for the filters of the 1st series (upper row, i.e. Figures 4.23a to 4.23c) and 2nd series (lower row, i.e. Figures 4.23d to 4.23f).

[112.5°;247.5°]) led to front-back confusions, e.g. back-to-front reversals, of in most 39.67%. In the 1st series, both stimuli led to a front-back confusion rate of 31.83%. The stimuli didn't show an effect on the front-back confusion rate (ANOVA p-value: $\gg 0.05$).

4.2.5.5.2 Left-right confusion Left-right confusions were obtained in 1.00% (1st series) and 3.69% (2nd series).

In the 1st, resp. 2nd series, the left-right confusion rate varied with the filters between 0.42% and 2.08%, resp. 0.25% and 12.37%, c.f. asterisks in Figure 4.23b, resp. Figure 4.23e. According to a standard ANOVA, the differences among the filters of the 1st series were not significant (p-value: $\gg 0.05$) while the differences among the filters of the 2nd series were significant (p-value: $\ll 0.001$). A pairwise ANOVA on the data of the 2nd series showed that filter N° 14s led to significant different results than the remaining four filters (pairwise p-values: $\ll 0.001$), while pairs of filters of the group {N° 1, N° 5, N° 6, N° 14d} did not show significant different results (pairwise p-values: > 0.05).

The azimuth angle ϕ_t had a significant effect on the left-right confusion rate in the 2nd series (p-value: < 0.001) but not in the 1st series (p-value: > 0.05). In the 2nd series, we obtained that the left-right confusion rate at median positions is up to 4.34% higher than at lateral positions of the same quadrant. Regarding only filter N° 14s, the left-right confusion rate at median positions was 14.77% and decreased for lateral positions to 7.95%.

4.2.5.5.3 Mixed confusion We obtained a mixed-confusion rate of 0.33%, resp. 3.72% in the 1st, resp. 2nd series.

The averaged, mixed confusion rates per filter ranged the intervals [0.0%;0.83%] (1st series) and [0.42%;12.36%] (2nd series). Based on a standard ANOVA, no significant difference was obtained for the data of the 1st series (p-value: > 0.05). In contrast, with the data of 2nd series we obtained a significant difference between the filters (p-value: $\ll 0.001$). With a pairwise standard ANOVA we identified that filter N° 14s led to significant different mixed confusion rates than the remaining filters (pairwise p-values: $\ll 0.001$). Pairs of filters, which are comprising filter N° 1, N° 5, N° 6, and N° 14d, did not show significant different effects (pairwise p-values: > 0.05).

The effect of the azimuth angle ϕ_t was not significant in the 1st series (p-value: > 0.05) but was significant in the 2nd series (p-value: < 0.001). In the 1st series, a mixed confusion rate of 0.50%,

resp. 0.17% was obtained at $\phi_t = 67.5^\circ$, resp. $\phi_t = 247.5^\circ$. Data of the 2nd series showed that within a quadrant the mixed-confusion rate increased from 2.75% at lateral positions to 5.17% at median positions.

4.2.5.6 Interactions

4.2.5.6.1 Angular error & Response time In Figures 4.18c, 4.18d and 4.20b we identified at the beginning of both series matching intervals where subjects showed long response times and large unsigned angular errors. Figures 4.24a and 4.24b show the trend of the unsigned angular error $|\epsilon|$ over the response time t_r in a 2D plot. Each dot represents one step on the scale of the test progress (coded by the dot's color) and indicates the corresponding $|\epsilon|$ and t_r , averaged over all subjects. The first, resp. last stimuli of a series is labeled by the letter S (start), resp. E (end). In the 2D plots, we easily recognize the random oscillation of $|\epsilon|$ and the convergence of t_r with increasing test progress, as it was already identified in Sections 4.2.5.1 and 4.2.5.3. The PCC of 0.40, resp. 0.15 for the 1st, resp. 2nd series showed that $|\epsilon|$ and t_r are not linearly correlated. This confirms that subjects showed only adaptation but no learning during the test.

Figures 4.24c and 4.24d were obtained by applying the same style of 2D plot but with respect to the rendering filters. Now each dot represents one rendering filter, indicating the corresponding value of $|\epsilon|$ and t_r . It can be noticed the filters of the 1st series are densely spread around the average unsigned angular error of 16.88° and the average response time of 2.37s, c.f. Figure 4.24c. These filters are not linearly correlated in the $|\epsilon|$ - t_r -plane which was supported by a PCC of -0.051 . Filters of the 2nd series, in particular filter N° 14d and N° 14s, led to a much larger variation around the average unsigned angular error of 29.36° and average response time 2.40s, c.f. Figure 4.24d. The corresponding PCC equaled to 0.75, indicating a certain correction between $|\epsilon|$ and t_r . This value was mostly caused by filters N° 14s and N° 14s and the small number of five data points. Nevertheless, subjects responded less intuitively (increased response time) and the provided directional cues were more ambiguous (increased absolute angular error) using the ambisonic method than the discrete method.

4.2.5.6.2 Front-back confusion rate & Hearing threshold The frequency range between 1 kHz and 10 kHz plays an important role for front-back discrimination, c.f. Blauert bands in Table 1.1. Hence

4.2. LISTENING TEST

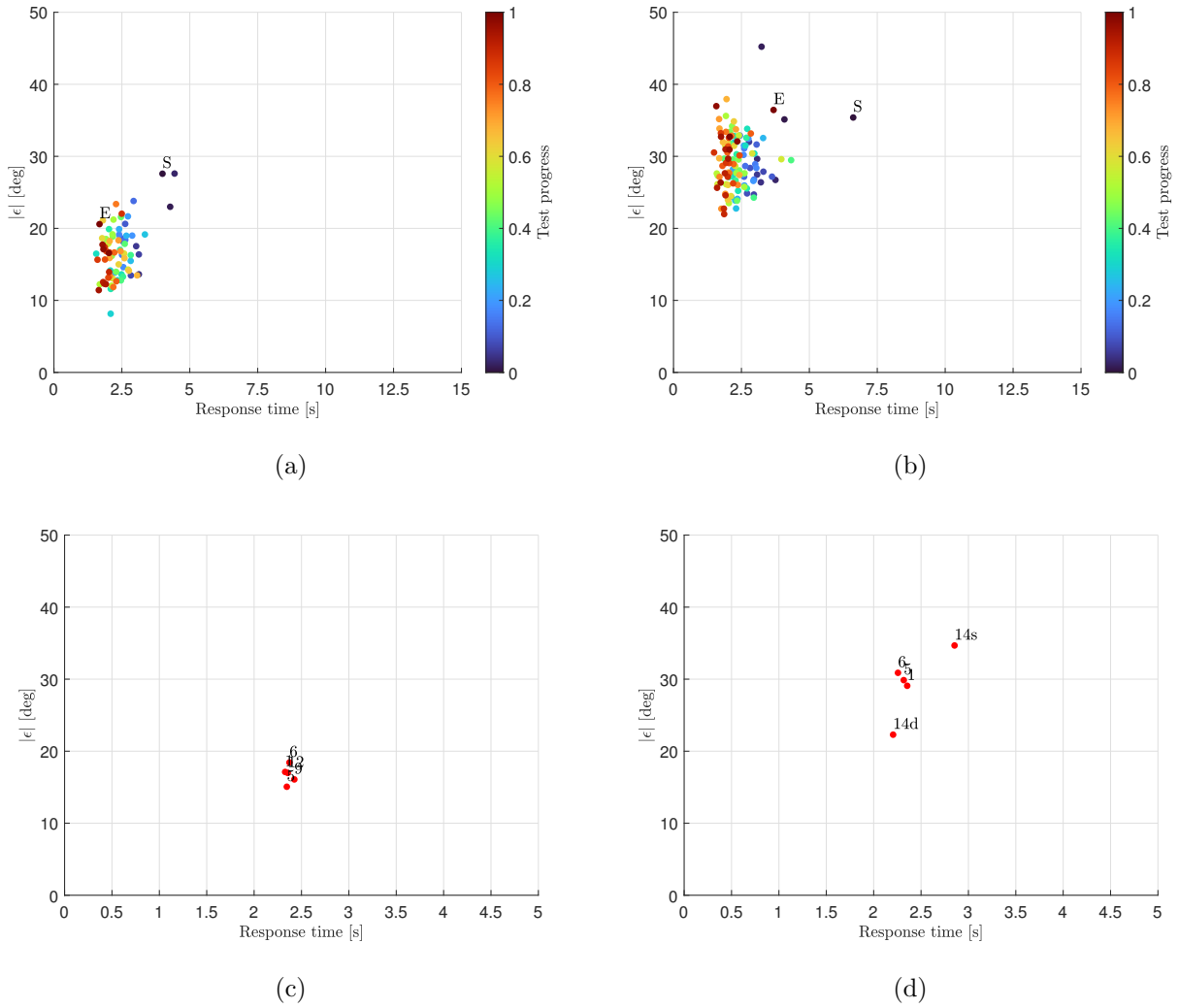


Figure 4.24: Unsigned angular errors $|\epsilon|$ over the response times of the 1st (Figures 4.24a and 4.24c) and 2nd (Figures 4.24b and 4.24d) series. In Figures 4.24a and 4.24b the dots represent one step on the test progress scale, c.f. trend along the test progress. The markers “S” and “E” highlight the start (test progress = 0) and end (test progress = 1) of the test. In Figures 4.24c and 4.24d the dots represent the data of the rendering filters.

it is of justified interest if the measured hearing threshold of the subjects in this frequency range and the observed front-back confusion rate correlated. Figure 4.25 visualizes the front-back confusion rates over the hearing thresholds at 1 kHz, 2 kHz, 4 kHz, and 8 kHz. Each dot represents the averaged data of one subject. By visual inspection, we notice in both series that the spread of the data increases with increasing frequency and the generally reduced front-back confusion rate in the 1st series compared to the 2nd series. In contrast the inter-subject variation of the front-back confusion rate was much larger in the 1st than in the 2nd series. Subject N° 17, who participated in both series, indicated prior to the audiometry that an audiologist had diagnosed a major increase in its hearing threshold which can be recognized in Figure 4.25 across all frequencies. The PCC between the front-back confusion rate and the hearing threshold for frequencies above 500 Hz ranged between -0.60 and -0.21 (1st series) and between 0.08 and 0.18 (2nd series). So, there was no linear correlation between the hearing threshold and the front-back confusion rates. Subjects with an increased hearing threshold over the entire frequency range, e.g. subject N° 17, perceived the stimuli quieter but with nearly the original spectral cues as all frequencies were dampened equally. Subjects with a narrow shift in hearing thresholds, e.g. subject N° 27, N° 30, N° 34, and N° 37, learned to compensate the loss by evaluating spectral cues which they are sufficiently sensitive to.

4.2.5.6.3 Front-back confusion rate & Series A possible effect of the series on the front-back confusion rate was examined by confronting data of both series in Figure 4.26. Only data sets where the directions of incidence, filters, and audio samples were matching between the series were compared. Therefore, we exclusively considered data of the 1st and 2nd series where the directions of incidence $\phi_t \in \{67.5^\circ, 247.5^\circ\}$, the filters N° 1, N° 5, and N° 6, and the male speech audio sample were used. Averaging the front-back confusion rate over these data sets of the 1st and 2nd series resulted in 45.83% and 44.19%. We notice the overlapping notches in Figure 4.26, i.e. no effect of the series on front-back confusion rates, which was confirmed by a Kruskal-Wallis test (p-value: $\gg 0.05$).

4.2.6 Discussion

The increased response time during the 2nd series seems to have been caused by the reduced number of 120 stimuli, compared to 208 stimuli during the 1st series. The progress bar of the user interface fills up faster during the 2nd series than during the 1st series, so subjects take little more time for

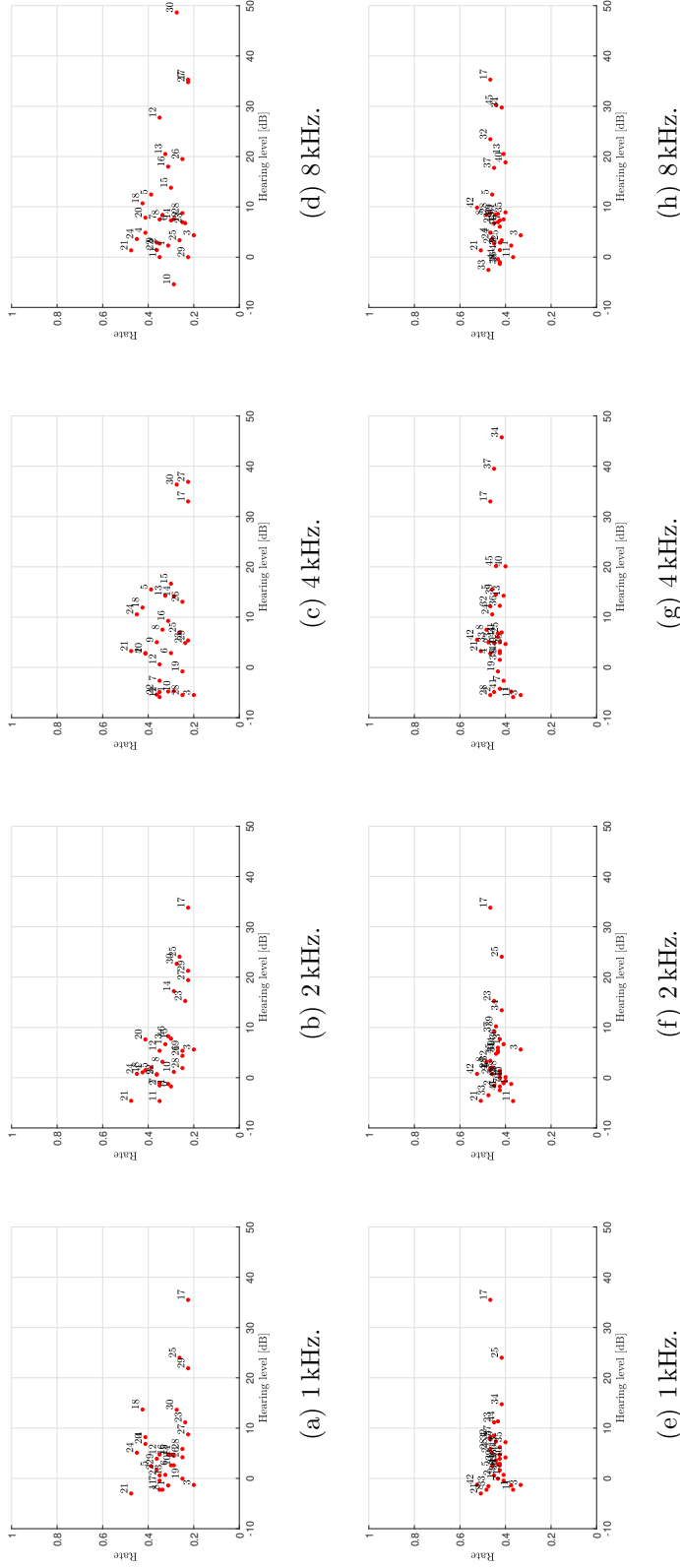


Figure 4.25: Front-back confusion rate over the hearing level at frequencies ≥ 1 kHz. The dots in Figures 4.25a to 4.25d (upper row), resp. Figures 4.25e to 4.25h (lower row) represent the average values for each subject of the 1st, resp. 2nd series.

4.2. LISTENING TEST

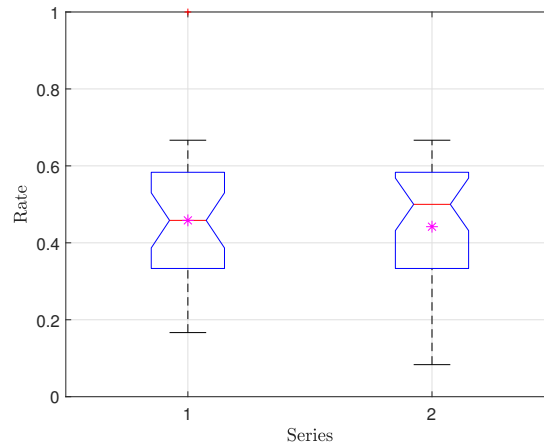


Figure 4.26: Box plot with mean values (asterisks) of the front-back confusion rates during the 1st and 2nd series. The graph is based on data where the filter, audio sample, and direction are the same for both series.

responding. The groups of rendering filters {N° 1, N° 5, N° 6, N° 14d} and {N° 14s} corresponds to the discrete and ambisonic processing method of the 2nd series. Stimuli which were filtered with the same HRTF but originate of different rendering methods, i.e. filter N° 14s and N° 14d, led to increased response times with the ambisonic method, i.e. filter N° 14s, than the discrete method, i.e. filter N° 14d. It seems that the ambisonic method leads to stimuli with less intuitive directional information, requiring the subjects to reflect about the spectral information and actively take decisions.

The decreasing response time and unsigned angular error indicate that subjects familiarize with the setup and the user interface mainly within the first 3% of the stimuli, i.e. test progress. The decrease in response time is also related to the monotonous task and the lack of re-attracting the attention of the subjects during the test. Subjects adapted to the test in the beginning of the 1st series. As most of the subjects participated in both series, most of them remembered quickly the procedure in the beginning of the 2nd series. This observation is very common as adaption and learning generally require more trials and time than remembering already learned patterns. The constant correct response rate along the test progress showed that the localization performance was not influenced by possible training effects.

Male speech might be more familiar to the subjects, and they have might perceived it as less disturbing than the cicada sound. the cicada sound resembled to noise which is suspected to be disturbing or annoying. Further, the sound level of the cicada sound sample was not realistic as

4.2. LISTENING TEST

subjects were allowed to individually set a pleasant sound level. Additionally, subjects were locals of the tripoint region Germany, France, and Switzerland, where the cicada is rare. It is more likely that subjects heard it on trips. Nevertheless, it seems that sound localization performance, regarding the angular error and correct responses rate, is significantly enhanced with an all-day sound than an unusual sound.

The high spread of the responded angle ϕ_r in the confusion matrices indicates that the directional cues were ambiguous for the subjects. Consequently, they had difficulty distinguishing between left-right and front-back. Subjects correctly localized more sounds with the discrete method than the ambisonic method. In the ambisonic processing, the virtual loudspeaker which is diagonally opposite of the quadrant where the sound source is located, emits a 180° phase shifted signal with a small amplitude. Except for $\phi_t \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, all four virtual loudspeakers emit acoustic signals for all sound directions. The phase shifted signal of the diagonal opposing loudspeaker augments the energy in the contralateral headphone signal and hence, increases the ambiguity about left-right discrimination. In contrast, the discrete rendering method introduces robust left-right cues by applying the ITD. The energy contribution of the diagonal opposing virtual loudspeaker also reduces the effect of the rendering filter. As the frequency responses of the rendering filters were generally symmetric for frontal and back sounds, the down mixing of the filtered, virtual loudspeaker signals led to a reduction of the previously introduced front-back cues.

The positions of the local maxima in Figure 4.20a reflect the response behavior of the subjects which we identified in the confusion matrices in Figure 4.19. The most frequent signed angular error is $\epsilon \in \{0^\circ, \pm 22.5^\circ\}$. This is caused by the false localization of the lateral positions $\phi_t \in \{67.5^\circ, 247.5^\circ\}$ at $\phi_r \in \{90^\circ, 270^\circ\}$ and the correct localization of the median positions $\phi_t \in \{0^\circ, 180^\circ\}$ at $\phi_r \in \{0^\circ, 180^\circ\}$. Subjects of the 2nd series tend to respond with a signed angular error of $\epsilon \in \{\pm 22.5^\circ, \pm 45^\circ, \pm 67.5^\circ\}$, which is in line with the previous findings that subjects localize falsely most of the stimuli at lateral positions. The location of the local minima at odd multiples of 11.25° and the distinct prominence of the local maxima show that subjects localized the stimuli within a finite set of discrete positions. Further, they managed to map precisely the space of the VAE to the circular, acoustic horizon of the user interface.

Our observation of the unsigned angular error $|\epsilon|$ in relation to the tested azimuth angle ϕ_t contributed an additional trend to what has been already published. We obtained a large unsigned angu-

lar error at median positions and decreasing unsigned angular errors for lateral positions. Contrarily, McAnally et al. reported a constant angular error over the entire horizontal plane [197], Makous et al. reported an angular error that increased with ϕ_t [198], while Pulkki et al. reported that the angular error reached maxima, resp. minima at lateral, resp. median positions [199]. A major difference to our work is that our subjects did not listen with their natural hearing, nor with their individual cues. In addition, the previously discussed problems with left-right cues in the ambisonic method cause large angular errors.

By definition, the calculation of the angular error projects all angular positions to the frontal hemisphere. This limits the maximum value of the angular error to

$$\begin{cases} \phi'_t + 90^\circ & , \text{ for } \phi'_t \leq 90^\circ \\ (360^\circ - \phi'_t) + 90^\circ & , \text{ for } \phi'_t \geq 270^\circ \end{cases} \quad (4.14)$$

with ϕ'_t denoting the projection of ϕ_t into the frontal hemisphere, c.f. Equation (2.7). Therefore, for the median and lateral positions, the maximum possible values are 90° and 180° , explaining the varying intervals in which the outliers were observed in Figure 4.21.

The test was designed as unforced choice test which is crucial for the accuracy of the responses and the different confusion rates. Responded angles that are close to the frontal or median plane were very likely mistakenly placed in the adjacent quadrant. Therefore, the confusion rates were highly influenced by the accuracy of the pointer position. The fact that the subjects most frequently responded with either a lateral or a median position, regardless of ϕ_t , reinforced this effect.

Stimuli where ϕ_t is located on the frontal plane, median plane, or on one of both planes is excluded from determining the front-back, left-right, and mixed confusion rates, c.f. Section 2.3.3.5. Hence, the front-back, left-right, and mixed confusion rates does not represent all data. This affects the left-right and mixed confusion rates of the 1st series, but not their front-back confusion rates. The left-right and mixed confusion rates were calculated on only 50% of the stimuli, i.e. where $\phi_t \in \{67.5^\circ, 247.5^\circ\}$. The 2nd series is not affected because all stimuli were located off the frontal and median planes. The little number of left-right and mixed confusions in the 1st series, was caused by the extreme values of the left-right cues at the few considered positions, making left-right discrimination evident, c.f. near-zero means in Figures 4.23b and 4.23c.

The left-right cues introduced by the ITD in the discrete method (filter N° 14d) led to significantly improved left-right discrimination than the cues introduced by the gain differences of the virtual

4.3. CONCLUSION

loudspeakers, i.e. the ILD, in the ambisonic method (filter N° 14s). In the 1st that contained only stimuli of the ambisonic method, subjects seemed to correctly interpret the left-right cues. In B we list issues of recording the stimuli by ambisonics and how they reduce the level differences in the decoded loudspeaker signals. This in turn inevitably led to a reduced ILD in the stimuli of the ambisonic method. When all stimuli contained weak left-right cues, i.e. 1st series of the listening test, subjects showed high sensitivity. In contrast, when presenting stimuli with strong and weak left-right cues, i.e. 2nd series of the listening test, the strong cues reduced the sensitivity such that weak cues were falsely interpreted. The subjective judgments about sound directions appear to be relative judgments between the provided cues in each series, and the results depend strongly on the set of provided stimuli.

The 1st series suffers the most from the similarities between the rendering filters. The development of the rendering filters was based on the same HRTF, resulting in highly related spectral cues. Subjects did not perceive differences between the rendering filters. The enormous number of filters, their only slight differences in frequency responses, and the lack of training apparently overwhelms the subjects.

The noticeably reduced front-back confusion rate of filter N° 14s in 2nd series was contrasted by a greatly increased left-right and mixed confusion rate. Based on the discrete rendering, the simulated, non-individual HRTF (filter N° 14d), led to more front-back confusions than simply boosting certain frequency intervals (filter N° 1, N° 5, and N° 6). Hence, when imposing non-individual spectral cues, generalized spectral cues with a smooth frequency response seems to be preferable rather than detailed spectral cues with a fine structured frequency response. Fewest front-back confusions were obtained neither with the highly generalized filter N° 1 nor with the narrow-peaked filter N° 6. Filter N° 5 seems to have been a good compromise between generalization and detail. Hence, the generalized cues must be designed with care and not too generic or have too narrow details. Additionally, filter N° 5 boosted high (≥ 1 kHz) and low (< 1 kHz) frequencies, while filter N° 1 and N° 6 boosted or attenuated high frequencies only (≥ 1 kHz).

4.3 Conclusion

Having started with the analysis of the spectral front-back differences based on simulated and measured HRTF of dummy heads, we identified the frequency range between 2 kHz and 8 kHz to contribute the most to spectral cues in the horizontal plane. In this frequency range the front-back

4.3. CONCLUSION

difference showed a wide band with medium gain, superposed by a narrow peak. With increasing azimuth angle the bandwidth of the wide band reduced, while the narrow peak first increased and then decreased.

Furthermore, 2nd-order IIR peak filters were used to develop a model representing the identified spectral front-back difference. The model was controlled by the horizontal angle ϕ and returned the filter coefficient of a corresponding 10th order IIR filter. The developed model had an RMSE < 1.7 dB and thus approximated very precisely the initially measured and simulated HRTF. The model provides a small, handy tool for obtaining a HRTF based filter without the need to store the entire, in general large memory demanding, HRTF.

Finally, a listening test was conducted in a headphone based VAE to evaluate the subjective sound localization performance. The stimuli were generated by ambisonic and discrete rendering methods. In both methods the spectral cues were applied to the signals by digital filters. With a large set of digital filters, different spectral cues were tested, including High-Shelf Filter (HSF), band boosting, and the previously developed HRTF model. The test was designed as unforced choice test and disposed of similar user interface to the localization test in Paragraph 1.4.2.2.4 and was split in two series of 208 (1st series) and 120 (2nd series) stimuli.

The obtained localization performance was identified to be influenced on one hand by the design of the test and on the other hand by the deployed test parameters. The unforced choice design has the greatest effect on the confusion rates for responded directions near the frontal or median plane. Due to a limited positioning accuracy of the response cursor, it easily happened that the responded angle fell into the neighboring quadrant. A forced choice test would reduce the number of front-back, left-right, and mixed confusions by simply canceling out confusions which are due to the inaccurate cursor positioning. Conducting the headphone based VAE in an office pod did not allow subjects to notice real world objects around them, which they could have identify as sound source. The absence of visual cues reduced the localization performance and made subjects to fall back on their daily experience, telling them that invisible sound sources are mostly located behind them. Coupling the VAE with a virtual visual environment, both simultaneously presented by a virtual reality headset, subjects are led to believe that loudspeakers or possible sound sources are placed all around them. The similarity of the filters and consequently the similarity of the spectral cues in the 1st series led to equal localization performance. Subjects did not perceive spectral cues altering between frontal and back

4.3. CONCLUSION

incidence, nor between the filters. In the 2nd series, the discrete method led to increased localization performance over the ambisonic method. We identified reduced ILD cues with the ambisonic method but also the issues related to ambisonic measurements as identified in Appendix B. From the filters of the discrete method, it turned out that in terms of front-back discrimination subjects preferred generalized frequency responses than highly detailed, non-individual HRTF. It seems that the detailed, non-individual HRTF was in concurrence with the detailed, individual HRTF of the subject, making the brain to match unsuccessfully the non-individual cues to its known, individual cues. In contrast, generalized frequency responses were recognized by the brain as such, making front-back discrimination more reliable.

In this chapter we showed how directional cues can be applied to an audio signal. We have identified low order filters to provide the most comprehensible front-back cues. The prototypes in the following chapter will use filter N° 1, i.e. the 2nd-order, IIR, 3 kHz HSF. This filter showed good localization results and is highly suitable for low power, embedded systems. Additionally, despite the trend towards low order filters, we will equip one prototype with KEMAR's HRTF that resembles to filter N° 14d but with a smoother frequency response. Knowing how to reconstruct the directional information under an HPD we will focus in the following chapter on how to capture the surrounding sound field.

4.3. CONCLUSION

Chapter 5

HPD prototypes

Content

5.1	Design aims	152
5.2	HPD Prototypes	152
5.2.1	Prototype A	153
5.2.2	Prototype B	159
5.2.3	Prototype C	160
5.2.4	Prototype D	163
5.3	Spectral cues	163
5.4	Localization performance	168
5.4.1	Stimuli	169
5.4.2	Setup	169
5.4.3	Procedure	171
5.4.4	Participants	172
5.4.5	Results	172
5.4.6	Discussion	180
5.5	Conclusion	183

The focus of this chapter is set on the development of improved HPDs regarding localization performance. Different designs are presented, and the assembled prototype were evaluated within a listening test. The experience gained from the listening test in the previous chapter is used for the setup of the upcoming listening test.

5.1 Design aims

The aims of the earmuff prototypes are the reconstruction of meaningful, spectral cues and enhancing sound localization with respect to the commercially available *Z-Tactical ZSORDIN* headset, c.f. HPD P2 in Figure 2.2. The prototypes are about gaining initial basic knowledge about the feasibility and functionality of the individual designs, but not about developing finished devices. To keep the setups simple we use post-processing and equipped the prototypes only with the most necessary electronic components, e.g. microphones. Therefore, they were not fully functional nor real time HPDs. This post processing approach facilitates the setups and the signal evaluation. Consequently, the subjective listening test requires a design where subjects listen to prerecorded stimuli instead of actually wearing the prototypes. The prototypes are evaluated regarding their usability, the recording techniques, i.e. the microphone configurations, and the localization performance.

Active talk through HPDs record surrounding sounds with external microphones and reproduce the recorded signals by internal loudspeakers. In order to provide improved sound localization, the prototypes need to provide meaningful spectral cues that code the directional information. These cues had to be introduced either before or after recording. Introducing them prior to recording implies that they have to part of the acoustic wave before it reaches the microphones. Introducing the cues after recording requires appropriate filtering of the recorded signals.

5.2 HPD Prototypes

We examined four HPD prototypes which were planned and assembled as acoustic sensors. The prototypes were not equipped with electronics, except the mounted external microphones. These prototypes allow to record signals, while any subsequent signal processing and reproduction were done offline on a PC. Even though the signal processing was performed on a standard PC, high importance was brought to its simplicity in order to guarantee viability on lightweight and portable embedded

HPD systems. For each prototype one pair of output signals was calculated which can be reproduced by any stereo headphones. These output signals correspond to those signals which will be reproduced under the prototypes, once the prototypes will be equipped with loudspeakers in the future.

The design of the prototypes, especially the geometries of the shells, was inspired by the design of HPD P2. The original design of HPD P2 was adapted to obtain a simplified (without any cut-outs for microphones or buttons), symmetric (with respect to the frontal plane) geometry. The prototypes must record the surrounding sounds and provide them to the user with the spatial information. Different microphone arrays were realized and tested across the prototypes. Binaural filtering of the audio signals, which provides spectral cues, was combined with stereo playback.

The shells and microphone supports of the prototypes, which were developed and constructed using numerical 3D models, were manufactured by additive technology using the *Stratasys* “Objet30” 3D printer. The deployed material was an opaque, rigid photopolymer (“VeroBlackPlus” by *Stratasys*) [200].

5.2.1 Prototype A

Prototype A consists of 6 cardioid, i.e. unidirectional, condenser microphones (*Kingstate* KEIG-4537TFL-N) which were mounted on the outside of the shells. The microphones were counterclockwise labeled from $M1$ to $M6$, starting with the front-left microphone, c.f. Figure 5.1. The 3D model of the shell with the mounting points for the microphones and the microphone support are shown in Figure 5.2. The symmetric shell was printed twice, the microphone support was printed six times. The microphones were clipped into microphone supports, each of which was attached to a mounting point. The microphone supports leave a 5 mm gap between the shell and the back of the microphones. The left and right shell were linked by a headband. The completed prototype with the microphone cabling is shown in Figure 5.3. On each shell the microphones were oriented in the three directions front, side, and back.

Figure 5.4a shows the measured, frequency dependent directivity characteristics of the microphone model used in free field without microphone support and shell. We notice sensitivity difference of 10 dB at 1 kHz between $\phi = 0^\circ$ (frontal direction) and $\phi = 180^\circ$ (back direction). The data provided by the manufacturer indicate a sensitivity difference of 20 dB at 1 kHz between frontal and back directions. The difference between the results of our measurements and the data of the manufacturer was caused

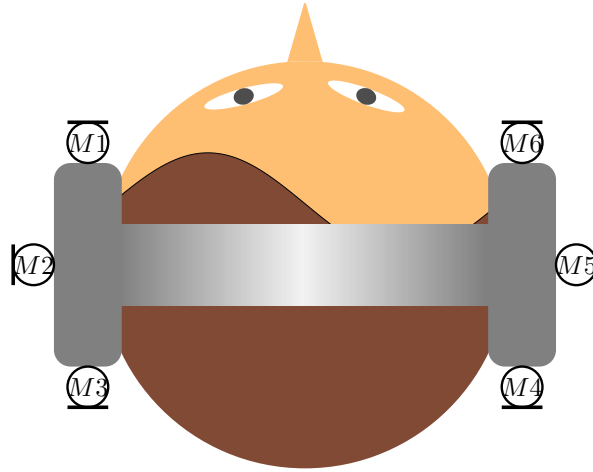


Figure 5.1: Sketch of Prototype A showing the positions of microphones $M1$ to $M6$ on the outside of the shell.

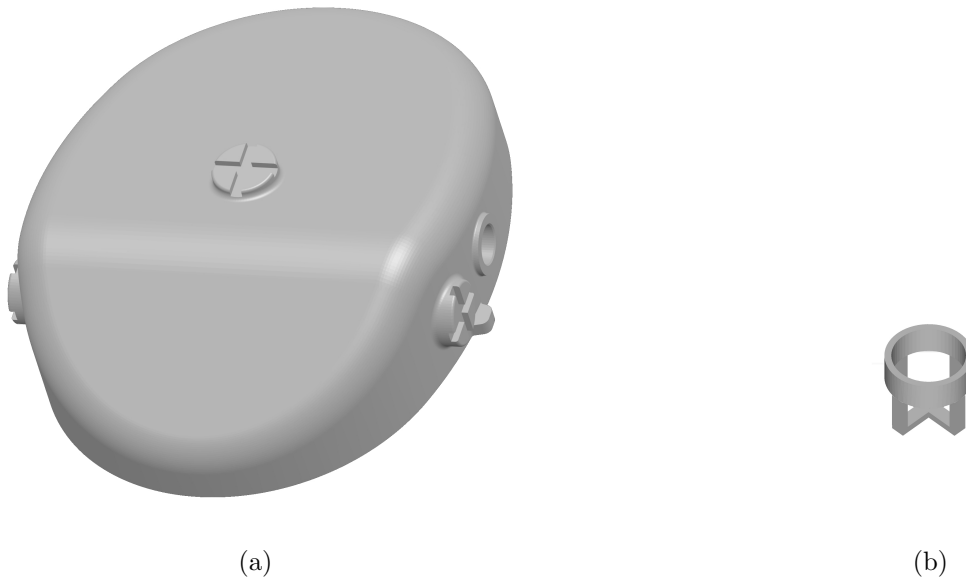


Figure 5.2: 3D models of the shells of Prototype A (Figure 5.2a) and the microphone support (Figure 5.2b). The mounting points for the microphone supports are the cross-like geometries on the outside of the shell in Figure 5.2a. The interactive 3D models are available when reading the electronic version of this document with an suitable PDF reader.

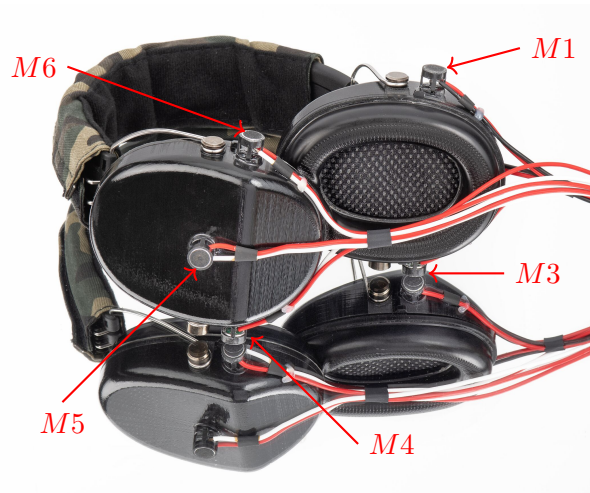


Figure 5.3: Prototype A with labeled microphones. Microphone M2 is located on the outside of the left shell facing away in the picture and is therefore neither visible nor labeled.

by the drift of the measurement equipment and the reflections at the microphone support we used during the measurements. Figure 5.4b shows the directivity diagram of the same microphone as used for Figure 5.4a but mounted with the microphone support on the shell of the prototype. We notice that the proximity between the microphone and the shell interacts with the directional characteristic of the microphone. The directivity trends towards omnidirectional characteristic, particular for frequencies below 2 kHz. The sensitivity difference between $\phi = 0^\circ$ (frontal direction) and $\phi = 180^\circ$ (back direction) reduces to 3.77 dB at 125 Hz and to 6.67 dB at 1 kHz. For frequencies above 1 kHz scattering and reflection causes multiple side lobes.

The directivity pattern of the assembled prototype was measured in the anechoic chamber with a realistic usage scenario by putting the prototype on dummy head Harry33. The measurement was done with pure tones and frequencies between 62.5 Hz and 16 kHz. They were spread by octaves and the angular resolution was set to 10° . The obtained directivity patterns of the front, side, and back microphones were symmetric between the left and right shell. For enhanced visibility Figure 5.5 shows the directivity patterns of the front, side, and back microphones, averaged between the left and right shell. It is noticed that the front and back microphones have approximately equal directivity patterns, each other symmetric to the frontal plane. The reduced cardioid characteristic of the microphones which was due to the proximity between the microphones and the shell, was recognized. The side microphones show symmetric characteristics with respect to the frontal plane. Side microphones

5.2. HPD PROTOTYPES

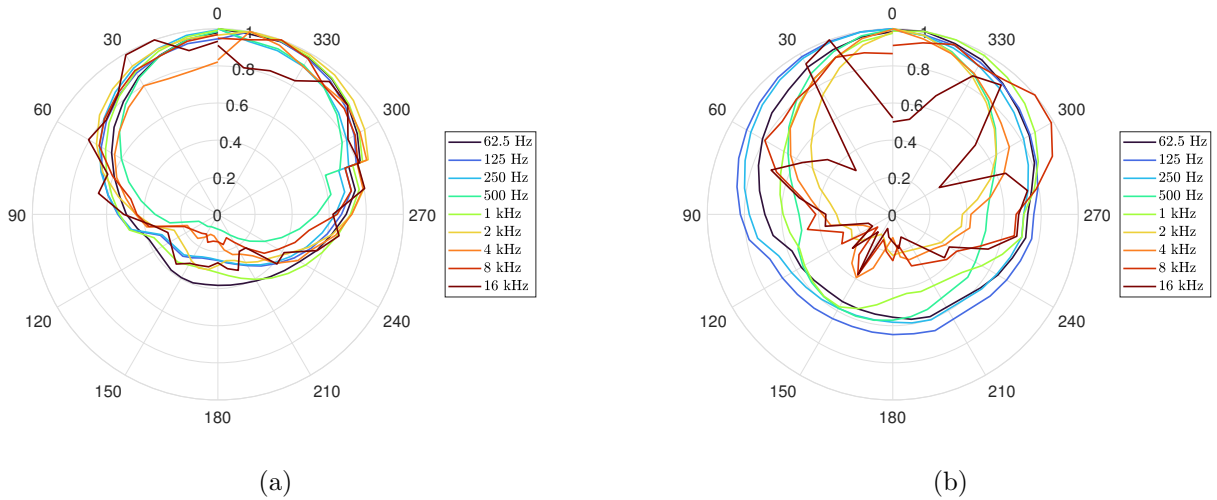


Figure 5.4: Frequency dependent directivity of the cardioid, i.e. unidirectional, microphone as measured in free field (Figure 5.4a) and mounted with distance of 5 mm on the shell (Figure 5.4b).

equally capture frontal and back sounds, blurring out directional cues in the subsequent filtering and summation stage. The sums of the directional patterns of the front and back microphones (FB) and the front, side, and back microphones (FSB) are plotted in Figure 5.6. Independently of the frequency, the FB configuration is slightly more sensitive for median directions and less sensitive for lateral directions as the FSB configuration. This difference is considered secondary to the blurring effect of the directional cues by the side microphone. It turned out that laterally positioned sounds were sufficiently captured by the front and back microphones. Hence, even though $M2$ and $M5$ were already mounted they are not going to be used anymore. It seems to be more promising, working from now on with the signals of $M1$, $M3$, $M4$, and $M6$, i.e. the front and back microphones, only.

Directional cues were applied to the equalized microphone signals x_{M1} , x_{M3} , x_{M4} , and x_{M6} by a 3 kHz HSF. According to [76], this filter approximates the front-back difference of HRTFs and enhances discrimination between frontal and back sound incidence. The gain of the HSF depends on the direction ϕ in which the corresponding microphone steers. Frank et al. propose for frontal directions, i.e. $M1$ and $M6$, resp. back directions, i.e. $M3$ and $M4$, a gain of $g_{HSF} = +6$ dB, resp. $g_{HSF} = -6$ dB [76]. High frequencies of frontal sounds are boosted while those of back sounds are diminished. In low frequencies the original magnitudes are kept. We use the implementation of 2nd order HSF from [194]. The output signals L_A , resp. R_A were obtained by summing the filtered signals

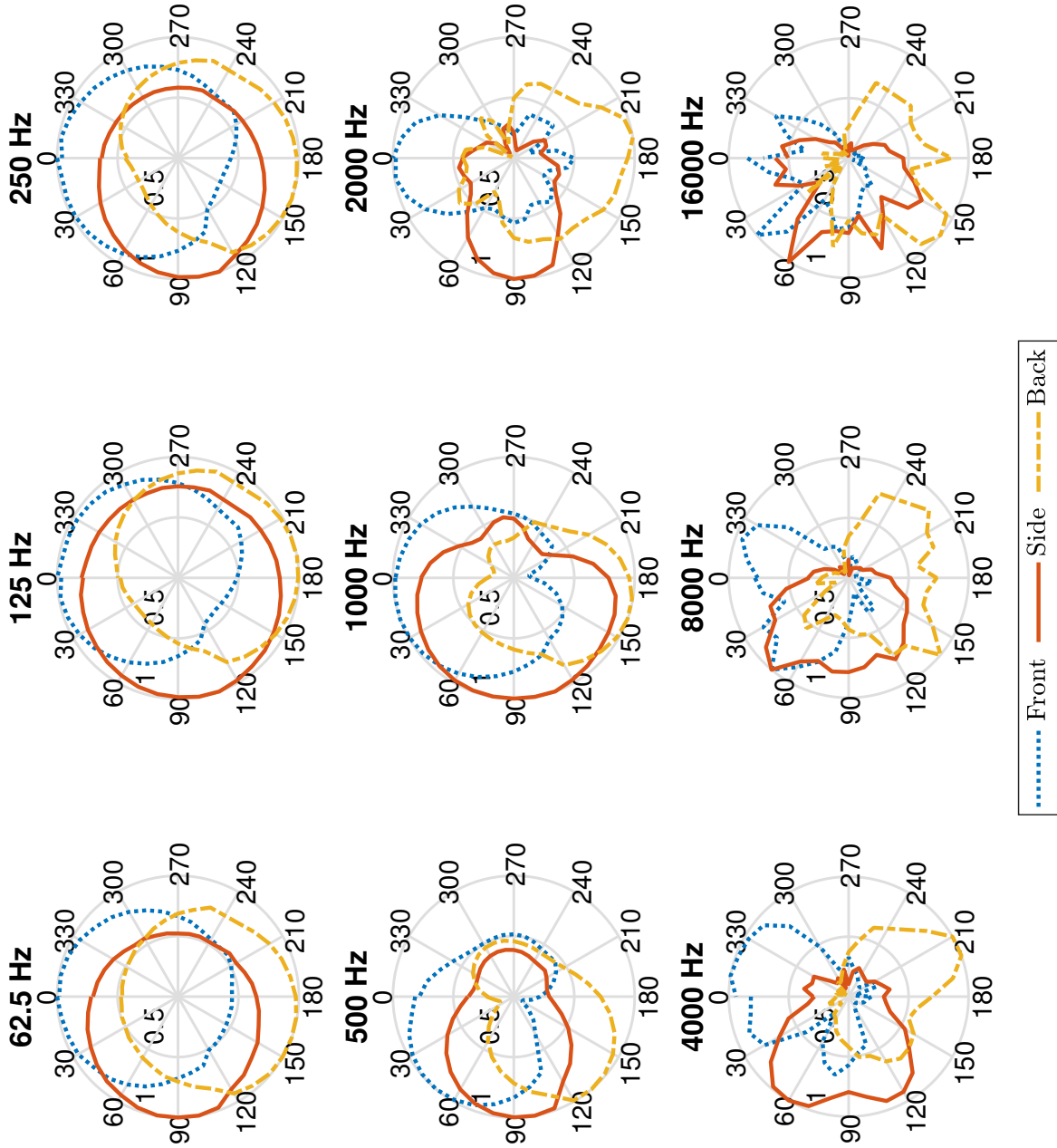


Figure 5.5: Directivity pattern of the front, side, and back microphones of Prototype A averaged between pairs of left and right microphones. The prototype was put on dummy head Harry33 for the measurements.

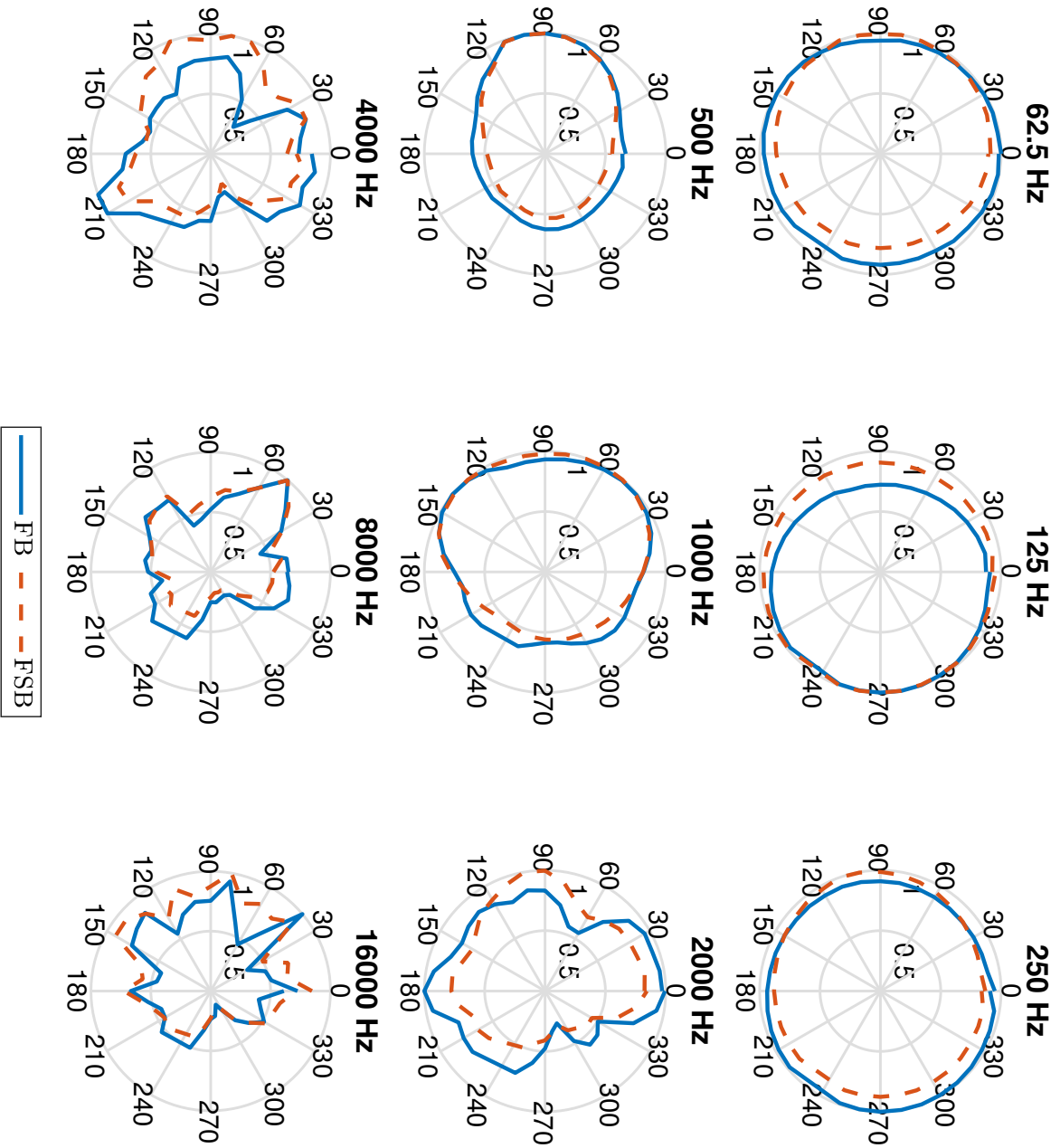


Figure 5.6: Directivity pattern of the Prototype A based on the front and back microphones (FB) and on the front, side, and back microphones (FSB). The prototype was put on dummy head Harry33 for the measurements.

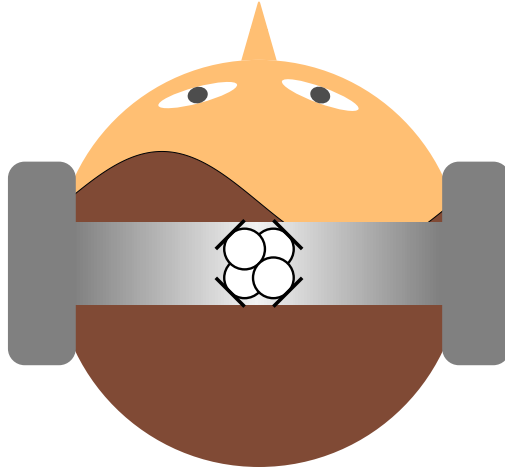


Figure 5.7: Sketch of Prototype B with the ambisonic microphone illustrated in the center of the headband.

of the left, resp. right microphones, c.f. Equation (5.1).

$$\begin{aligned} L_A &= HSF_{+6\text{ dB}} * x_{M1} + HSF_{-6\text{ dB}} * x_{M3} \\ R_A &= HSF_{-6\text{ dB}} * x_{M4} + HSF_{+6\text{ dB}} * x_{M6} \end{aligned} \quad (5.1)$$

5.2.2 Prototype B

Prototype B captures the sound filed around the listener with an ambisonic microphone, while the shells were only required for holding the HPD on the head, Figure 5.7. The shells were constructed symmetrically, c.f. 3D model in Figure 5.8, printed twice, and assembled with a headband. A 1st order ambisonic microphone (*Core Sound* “TetraMic”) was mounted to the headband, c.f. Figure 5.9.

The ambisonic A-format signals, i.e. the output signals of the four microphone capsules, were equalized, encoded to ambisonic B-format, and decoded for a squared setup of virtual loudspeakers by using the VVMic application by *VVAudio* [201]. The virtual loudspeakers $V1$ to $V4$ were virtually positioned between $\phi = 45^\circ$ and $\phi = 135^\circ$ with a spacing of 90° in space around the prototype or listener. Similar to Prototype A, the loudspeaker signals x_{V1} to x_{V4} were filtered with the 3 kHz HSF, which is low cost in terms of computational resource and approximates the front-back difference of HRTFs, c.f. Section 5.2.1. For the frontal signals x_{V1} and x_{V4} , resp. the back signals x_{V2} and x_{V3} the gain of the filter was set to +6 dB, resp. -6 dB [76]. The output signals L_B and R_B were obtained

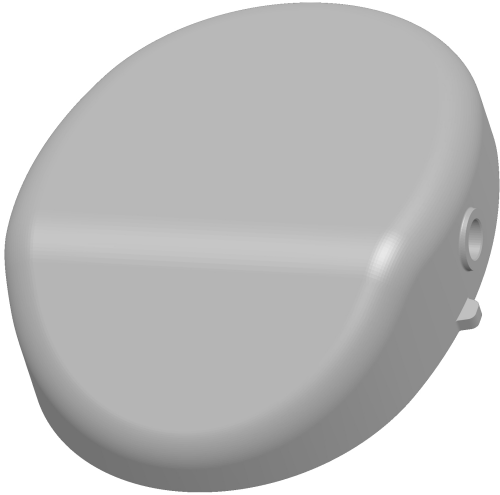


Figure 5.8: Symmetric 3D model of the shells of Prototype B. The interactive 3D model is available when reading the electronic version of this document with a suitable PDF reader.



Figure 5.9: Prototype B with the 1st order ambisonic microphone. *AmbMic* marks the tetrahedral arranged microphone capsules of the ambisonic microphone.

by linear combination of the filtered loudspeaker signals, c.f. Equation (5.2).

$$\begin{aligned}
 L_B &= HSF_{+6\text{dB}} * x_{V1} + HSF_{-6\text{dB}} * x_{V2} \\
 R_B &= HSF_{-6\text{dB}} * x_{V3} + HSF_{+6\text{dB}} * x_{V4}
 \end{aligned}
 \tag{5.2}$$

5.2.3 Prototype C

Prototype C aims to provide compatibility with ballistic helmets, and therefore its design was not based on HPD P2. The acoustic array consisting of 6 cardioid, i.e. unidirectional, condenser microphones (*Kingstate* KEIG4537TFL-N), was therefore mounted directly on the ballistic helmet, c.f. Figure 5.10. Gillett, Hengy, or Capin et al. examined similar approaches but with omnidirectional microphones and subsequent diffraction compensation [202, 203, 204]. To provide acoustic protection and reproduce the acoustic signals, users must wear appropriate earplugs or earmuffs. The microphones were mounted with the microphone supports, c.f. Figure 5.2b, on the outside of the helmet in a plane parallel to the horizontal plane, c.f. Figure 5.11. The microphones were distanced by 5 mm to the surface of the helmet. Microphone *M1* points to the front and microphones *M2* to *M6* follow

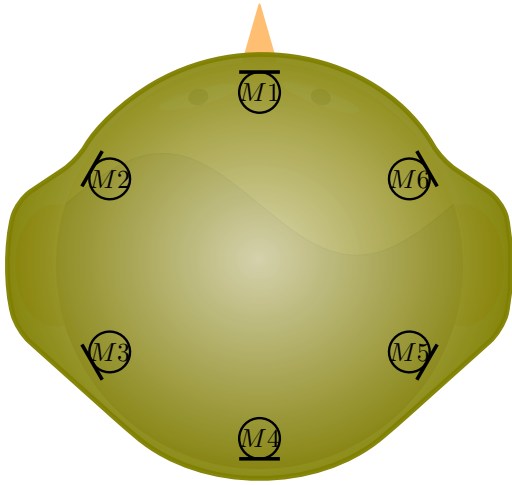


Figure 5.10: Sketch of Prototype C indicating the positions of microphones $M1$ to $M6$ on the helmet.



Figure 5.11: Prototype C with labeled microphones. Microphone $M3$ and $M4$ are not visible and therefore nor label.

counterclockwise, c.f. Figure 5.10. The angular distance between two neighboring microphones is 60° .

Figure 5.12 shows the directivity pattern of microphones $M1$ to $M4$ in octave bands between 62.5 Hz and 16 kHz. Data of microphones $M5$ and $M6$ are not shown in order to increase visibility. Their data were symmetrical to $M3$ and $M2$ with respect to the median plane. Diffraction around the helmet and the proximity between the microphones and the helmet cause the unidirectional characteristic of the microphones in free field to be altered towards omnidirectional characteristic when mounted on the helmet, also compare Figure 5.12 with Figure 5.4.

The equalized microphone signals x_{M1} to x_{M6} were filtered with KEMAR's HRTF $\mathcal{H}(\phi)$, where ϕ corresponds to the angular position of the microphones. The HRTF of the left ear, i.e. the one with the "normal"-sized pinna, was taken from the MIT database [140]. $\mathcal{H}(\phi)$ was implemented as a FIR filter of 128 points.

The signal x_{M1} was filtered with $\mathcal{H}(0^\circ)$, x_{M2} and x_{M6} were filtered with $\mathcal{H}(60^\circ)$, x_{M3} and x_{M5} were filtered with $\mathcal{H}(120^\circ)$, and x_{M4} was filtered with $\mathcal{H}(180^\circ)$. The output signals L_C and R_C were the weighted sums of the filtered signals. Microphones which are entirely positioned in the left, resp. right hemisphere exclusively contribute to the left L_C , resp. right R_C output signal. The corresponding microphones signals were therefore weighted with a factor of 1. Microphones which are located on the median plane contribute to both output signals, c.f. Equation (5.3). To prevent that these signals

5.2. HPD PROTOTYPES

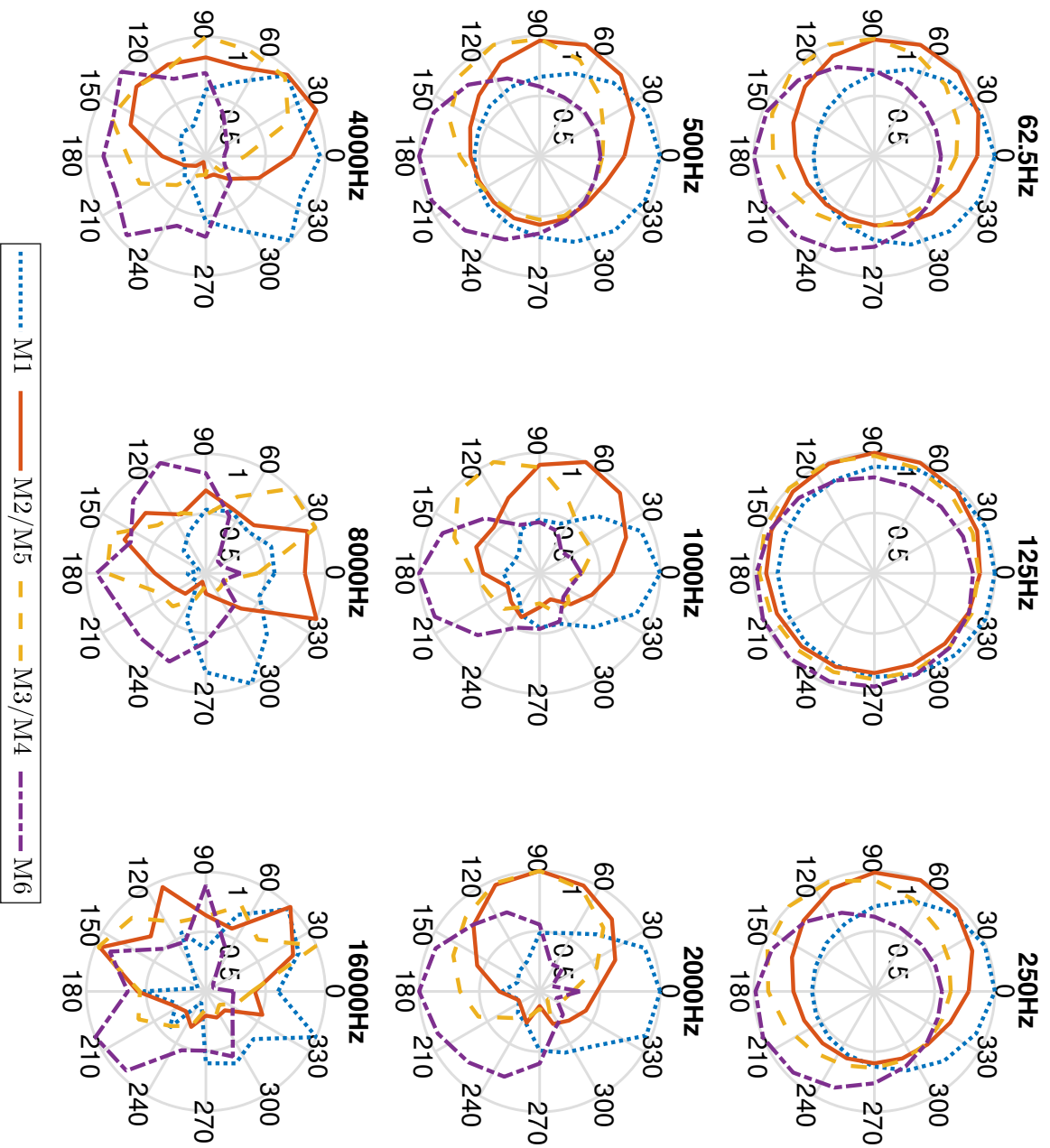


Figure 5.12: Directivity patterns of the mounted microphones of Prototype C. Due to symmetry along the median plane the pairs $M2$ and $M6$, resp. $M3$ and $M5$ are represented by the data of $M2$, resp. $M3$. The prototype was put on dummy head Harry33 for the measurements.

were perceived much louder than off median plane microphone signals, we applied a gain of $\frac{1}{2}$ to them.

$$\begin{aligned} L_C &= \frac{1}{2} \cdot \mathcal{H}(0^\circ) * x_{M1} + \mathcal{H}(60^\circ) * x_{M2} + \mathcal{H}(120^\circ) * x_{M3} + \frac{1}{2} \cdot \mathcal{H}(180^\circ) * x_{M4} \\ R_C &= \frac{1}{2} \cdot \mathcal{H}(0^\circ) * x_{M1} + \mathcal{H}(60^\circ) * x_{M6} + \mathcal{H}(120^\circ) * x_{M5} + \frac{1}{2} \cdot \mathcal{H}(180^\circ) * x_{M4} \end{aligned} \quad (5.3)$$

5.2.4 Prototype D

Prototype D was intended to introduce the spectral cues prior to recording by the geometry of its shells, similarly to the principle of the outer ear. Rubak et al. carved a concha-like cavity in the shell of an earmuff hearing protection but never provided localization performance [132] while Joubaud attached an outer ear-like geometry on top of the microphones of an active HPD [166]. Protruding geometries seemed not very useful, so we decided to pick up the idea from Rubak.

Instead of the triangular height profile, as used for the shells of Prototype A and Prototype B, a constant height profile was used for the shells of Prototype D. This allows to carve a cavity into the shell which follows the geometry of the simplified pinna simulator Type 3.4, c.f. Figure 5.14 [147]. A left and right shell was manufactured. Both were each equipped with an omnidirectional pressure field condenser microphone placed at the ear canal entrance point of the pinna simulator. Except of the directionality, the microphones characteristics were like those of *Kingstate* KEIG4537TFL-N. The shells were assembled by a headband, c.f. Figure 5.15. The stereo output signals L_D resp. R_D were the equalized microphone output signals of the left x_{M1} , resp. right x_{M2} microphone.

5.3 Spectral cues

Dummy head HATS33 was placed in the center of the circular loudspeaker array in the semi-anechoic chamber, c.f. Figure 5.16a, and exposed to white noise which was subsequently presented at frontal and back positions, i.e. $\phi \in \{0^\circ, 180^\circ\}$. The duration of the white noise at each position was 5 s. Recording was performed with a sampling frequency of 48 kHz, an audio interface (*MOTU* “8pre-es”) and a digital audio workstation (*Steinberg* “Cubase”). Six recording series were performed each simulating one of the following hearing conditions (HC):

HC0 HATS33 with its natural hearing, i.e. without any HPD nor prototype, c.f. Figure 5.16a. Output

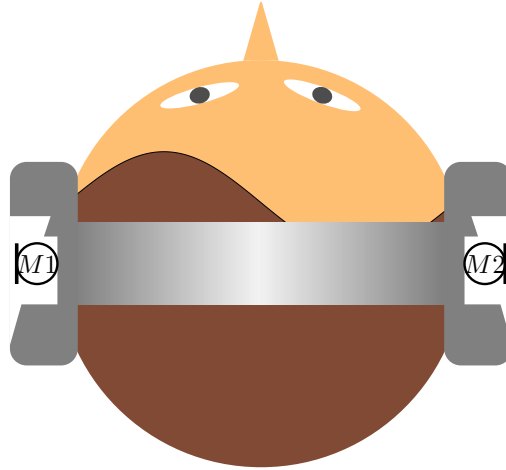


Figure 5.13: Sketch of Prototype D showing the cavities in the shells and the positions of the microphones $M1$ to $M2$.



Figure 5.14: 3D model of the left shell of Prototype D. The microphone is integrated in the cutout at the ear canal entrance point of the concha-like cavity. The interactive 3D model is available when reading the electronic version of this document with an suitable PDF reader.



Figure 5.15: Prototype D with labeled microphones in the concha-like geometry of the shell. The interior of the right shell is visible and reveals the back view of microphone $M2$.

5.3. SPECTRAL CUES

signals of HATS33 were recorded.

HC1 HATS33 wearing HPD P2, c.f. Figure 5.16b. Output signals of HATS33 were recorded.

HC2 HATS33 wearing Prototype A, c.f. Figure 5.16c. Signals of the microphones $M1$, $M3$, $M4$, and $M6$ of Prototype A were recorded.

HC3 HATS33 wearing Prototype B, c.f. Figure 5.16d. Signals of the ambisonic microphone of Prototype B were recorded.

HC4 HATS33 wearing Prototype C, c.f. Figure 5.16e. Signals of the microphones $M1$ to $M6$ of Prototype C were recorded.

HC5 HATS33 wearing Prototype D, c.f. Figure 5.16f. Signals of the microphones $M1$ and $M2$ of Prototype D were recorded.

For HC0, resp. HC1, the internal microphones of HATS33 were used for recording, providing the pair of stereo output signals L_{HC0} and R_{HC0} , resp. L_{HC1} and R_{HC1} . The microphone signals were not equalized to remove the Transfer Function of the Open Ear (TFOE). When these stimuli are listened through headphones, the signal arriving at the eardrum contains the individual TFOE of the listener and the TFOE of the dummy head. For HC2 to HC5, the dummy head was required to simulate a realistic use of the prototypes, but the microphones of the prototypes were used for recording. The recorded signals were processed accordingly to Sections 5.2.1 to 5.2.4. The signals L_{HC2} to L_{HC6} and R_{HC2} to R_{HC6} correspond to the output signals L_A to L_D and R_A to R_D .

The spectral front-back difference ΔH_{HCi} ($i \in [0, 5]$) of the pair of stereo output signals L_{HCi} and R_{HCi} was calculated and averaged for each hearing condition between the left and right signal, c.f. Equation (5.4). The results are plotted in Figure 5.17.

$$\Delta H_{HCi} = 20 \cdot \log_{10} \left(\left| \left(\prod_{\xi=\{L_{HCi}, R_{HCi}\}} \frac{\text{FFT}(\xi(\phi = 0^\circ))}{\text{FFT}(\xi(\phi = 180^\circ))} \right)^{\frac{1}{2}} \right| \right) \quad (5.4)$$

As seen in Figure 5.17, the weak signal to noise ratio, i.e. the noise floor of the measurement setup, causes high frequency (> 20 kHz) peaks and notches in ΔH_{HC0} to ΔH_{HC5} . Except ΔH_{HC5} , the curves shows a broad boost band in the frequency interval [3 kHz; 7 kHz] (HC0), [2 kHz; 7 kHz]



Figure 5.16: Figure 5.16a: Dummy head HATS33 placed in the semi-anechoic chamber, ready for HC0 measurements. According to this setup, the measurement setups for HC1 to HC5 are similar but with the corresponding HPD, resp. prototype put on the dummy head. Figures 5.16b to 5.16f: Close-up views of HATS33 wearing HPD P2 (Figure 5.16b) and Prototype A to D (Figures 5.16c to 5.16f).

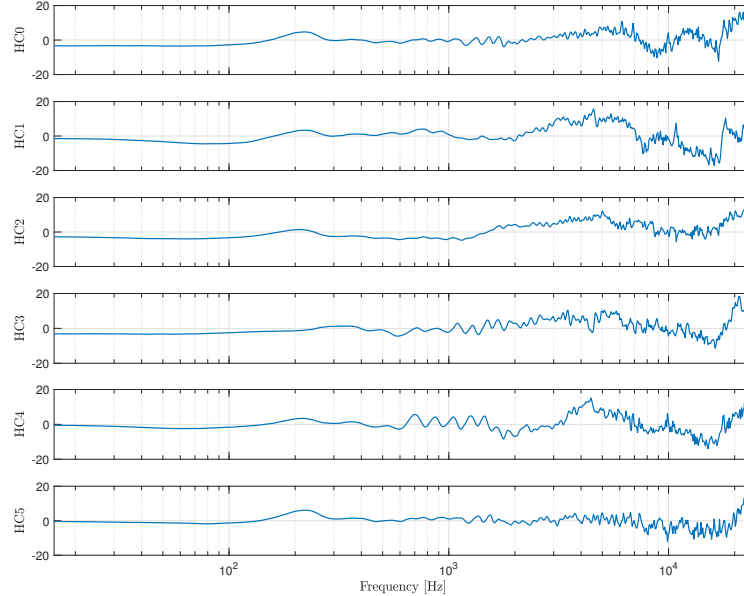


Figure 5.17: Magnitude of the spectral front-back difference ΔH between median sound source positions of $\phi = 0^\circ$ and $\phi = 180^\circ$ for the six hearing conditions HC0 to HC5. The curves are zero averaged.

(HC1), [1.5 kHz; 8.5 kHz] (HC2), [2.5 kHz; 6 kHz] (HC3), and [3.5 kHz; 7 kHz] (HC4). This broad boost band shows an additional peak on top which is least present for HC3 (10.48 dB), followed by HC0 (10.98 dB) and HC2 (12.26 dB) and most present for HC4, resp. HC1 (15.21 dB, resp. 15.62 dB). Even though this peak has individual prominence and width for each hearing condition, the boost band aligns with Blauert’s boosted band of frontal directions between 1.86 kHz and 7.03 kHz [40]. In contrast, such characteristics can not be obtained for ΔH_{HC5} . The corresponding curve varies closely around 0 dB on the entire frequency range. Only a small peak at 200 Hz and a slight attenuation in the high frequencies (> 10 kHz) are noticed.

The design of Prototype D appears to result in no front-back being measured in HC5. The membrane of the condenser microphone is located behind a centered, 2 mm circular opening in the front of the microphone housing. We define the center point of the microphone to represent the microphone membrane. In the cross-sectional view of the shell of Prototype D, c.f. Figure 5.18, α denotes the sector for direct line of sight with the center point of the microphone. α ranges from 66.46° to 126.76° for the left shell and from 233.24° to 293.54° for the right shell. Hence, it spans 60.30° slightly oriented to the back. Sound sources which are positioned within the sector α are in

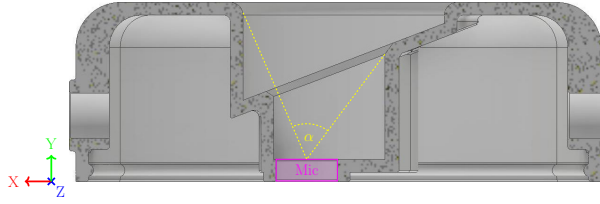


Figure 5.18: Cross-sectional view along the positive Z -axis of the left shell of Prototype D. The position of the microphone is marked by a magenta colored rectangle, c.f. “Mic”. α denotes the sector for direct line of sight with the center point of the microphone.

direct line of sight of the microphones. Sound emitted from these positions reach the microphones on direct, reflected, and diffracted sound paths. Sound sources which are positioned outside the sector α , including median positions, reach the microphone only on reflected and diffracted sound paths. The diffraction at the outer edges of the concha-like cavity, c.f. end points of the dotted lines in Figure 5.18, in combination with reflections result in similar spectral cues for frontal and back sound incidence. This canceled out any front-back differences and led to the flat ΔH_{HC5} in Figure 5.17. The cavity on the outside of the shell was expected to introduce spectral cues unique to each direction, similar to the principle of the outer ear. In contrast, the developed design for Prototype D did not introduce noticeable spectral differences between $\phi = 0^\circ$ and $\phi = 180^\circ$.

5.4 Localization performance

Since the prototypes were mounted as acoustic sensors, the localization performance must be assessed by an “offline” listening test, and cannot be assessed by a “live” localization test, as conducted in Chapter 2. The prototypes were used to record the stimuli. The signal processing was done offline, and the output signals were presented through stereo headphones to subjects. The output signals simulated the wearing of the prototypes. Subjective sound localization performance with the presented Prototypes A to D was compared to subjective sound localization performance without earmuffs and with HPD P2 (*Z-Tactical ZSORDIN* headset).

This localization test differs from the preceding listening and localization test concerning the tested direction ϕ_t of the stimuli. This test used dynamic stimuli positions, i.e. rotating stimuli, which lead in general to better localization performance than static stimuli positions [34]. We limit the range of rotation to the four quadrants left-front (LF), left-back (LB), right-back (RB) and right-front (RF).

We decided to use a forced choice design to eliminate the inaccuracies in the positioning of the cursor.

5.4.1 Stimuli

The stimuli were recorded in the semi-anechoic chamber with the same hardware setup as used for the measurements of the spectral cues in Section 5.3. With vector base amplitude panning [173] a sound sample was circularly moved on eight trajectories with constant radius of 1.10 m around the dummy head. The radius of 1.10 m equals the radius of the loudspeaker array. The start positions ϕ_S of the trajectories were located at $\phi_S \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$. End positions ϕ_E of the trajectories were located at $\phi_E = \phi_S \pm 90^\circ$. Hence, the sound sample swept counterclockwise and clockwise around the dummy head through the four quadrants on the eight trajectories, respectively. The angular speed of $\omega = 0.65 \text{ s}^{-1}$ was defined by the 90° arc through which the sound sample swept and the duration of the sound sample of $t = 2.41 \text{ s}$. The recording was sampled at 48 kHz and performed with all six hearing conditions HC0 to HC6, c.f. Section 5.3.

Two series of recordings were performed with two different sound samples. A natural, pleasant sound which subjects were generally familiar with, i.e. male speech, c.f. Figure 5.19a, and a sound which stimulates the human hearing system on the entire audible frequency range, i.e. pulsed white noise, c.f. Figure 5.19b, were used. As the pulsed white noise covers the entire audible frequency range, it was aimed to provide more spatial information than the male speech. The post processing, which was required for obtaining the output signals, was done as described in Section 5.3.

The total set of stimuli contained 96 stereo audio files (2 sound samples, 6 hearing conditions, 4 quadrants, 2 directions of rotation). Due to filtering and avoidance of temporal clipping, the recorded stimuli lasted 3.00 s.

5.4.2 Setup

Subjects were seated in an office pod (*SBS Silence Business Solutions* “Procyon Quatro”) and listen to stimuli through stereo headphones (*BeyerDynamics* “DT 770M”). The headphones were connected to a tablet computer that guided the listening test and served as user interface. The GUI provided 6 response buttons (one for each quadrant and each direction of rotation), a replay button, and a progress bar, c.f. Figure 5.20.

Subjects were asked to indicate the perceived quadrant and direction of rotation of each stimulus

5.4. LOCALIZATION PERFORMANCE

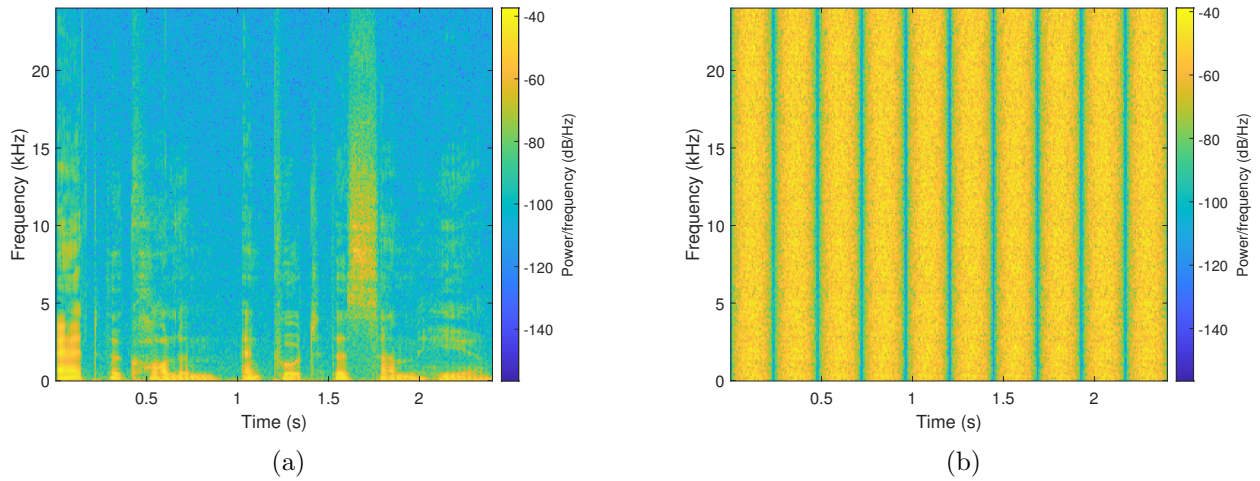


Figure 5.19: Spectrogram of the sound samples which were used for recording the moving stimuli. Male speech with limited energy in the high frequencies (Figure 5.19a) and pulsed white noise (Figure 5.19b).

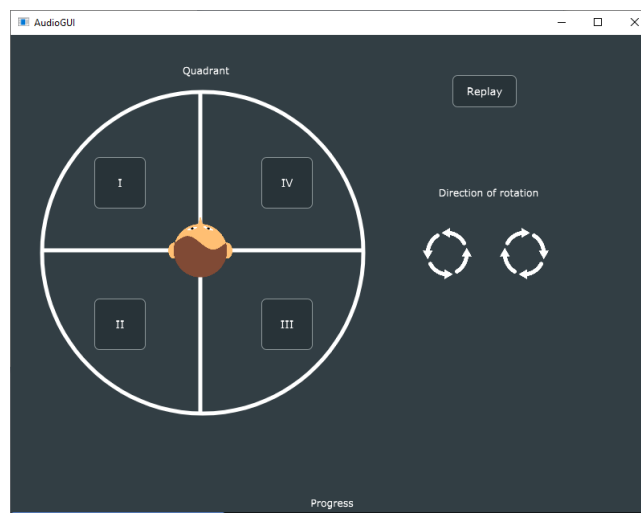


Figure 5.20: GUI for the listening test.

5.4. LOCALIZATION PERFORMANCE

by pressing the corresponding response buttons on the user interface. Subjects were not given the opportunity to change their responses afterwards. Once, subjects responded a quadrant and a direction, the next stimulus was automatically presented. Subjects had the possibility to listen the current stimulus a second time by pressing the replay button. This functionality was introduced because the number of stimuli to listen to was large and subjects might become inattentive during the test. We wanted subjects to have the opportunity to re-listen the current stimulus in order to interpret the spatial information, rather than responding arbitrarily. The replay button became inactive for the current stimulus as soon as one of the 6 response buttons was pressed.

5.4.3 Procedure

Both sounds samples were tested during the listening test, which therefore was split into two series. During the 1st series, subjects listened to stimuli containing either male speech or white noise. The other stimuli were presented during the 2nd series. The order was randomly chosen for each subject. Nevertheless, it was ensured that a balanced number of subjects started with the male speech sample and pulsed white noise sample. Each series contained 48 different stimuli (2 directions, 4 quadrants, 6 hearing conditions) and each stimulus was repeated 3 times, resulting in 144 stimuli per series. The two series were separated by at least 7 days.

The subjects were split into three groups. Each group was assigned with an individual order regarding the presentation of the stimuli. Subjects in group 0 listened to the 144 stimuli in random order. Subjects of group 1 and 2 listened the stimuli in subsets of 24 stimuli. Each subset comprised the 24 stimuli of one of the six hearing conditions. The order of the stimuli was random within the subsets. The order of the subsets was also random. Only subjects of group 2 were informed by a message on the tablet's screen that the upcoming stimuli originate of a different processing method, i.e. hearing condition subset.

Upon arrival subjects were introduced to the listening test. They were told that they will hear moving sounds through headphones and that their first impression counts about the quadrant in which they perceive the sound, but also the direction of rotation of the sound. They were instructed to use the playback button according to its purpose, i.e. not to overuse it and not to use it to confirm their decisions. Subjects in group 2 were told that different processing methods are tested, the stimuli are grouped by processing method, and that they are informed when the processing method changes.

5.4. LOCALIZATION PERFORMANCE

Subjects in group 1 and 0 were not told anything about the processing methods. After completing the series with the pulsed white noise, subjects were asked what they associated with the sound they just heard. After completion of the 2nd series, subjects were asked which of the two sound samples they would prefer in terms of least annoyance and easier localization of the sound during a hypothetical third series.

5.4.4 Participants

36 subjects participated in the listening test, 12 in each group. Subjects were between age 21 years and 54 years with a mean of 31.89 years. As subjects were randomly split into the three groups, the mean age per group was not constant ($\mu = 34.08$ years in group 0, $\mu = 28.42$ years in group 1, $\mu = 33.17$ years in group 2). Some of the subjects occasionally had participated in listening tests in the past, others not at all. None of them participated regularly in listening tests. The entire listening test ran for 18 days. 8 days passed in average between the 1st and 2nd series.

5.4.5 Results

A Kolmogorov–Smirnov test was performed to verify the null hypothesis that the data were following standard normal distribution. If the null hypothesis was rejected at the 5% level, a Kruskal-Wallis test was performed to check for significance, otherwise standard ANOVA was applied.

5.4.5.1 Response time

The average response time was 2.34 s with a standard deviation of 1.68 s. During the 1st, resp. 2nd series the response time was 2.57 s, resp. 2.11 s. According to a Kruskal-Wallis test, this difference was statistically significant (p-value < 0.05). The response time, averaged over all subjects, was 2.33 s for male speech stimuli and 2.36 s for pulsed noise stimuli. With a Kruskal-Wallis test it was proven that there was no effect of the type of stimuli on the response time (p-value: $\gg 0.05$). The response time, averaged over all subjects, decreased with increasing test progress with (inverse) exponential characteristics, c.f. Figure 5.21. With least squares were determined the model for the response time $t_r(x)$ to be

$$\hat{t}_r(x) = 1.98 + 1.86 \cdot e^{-5.07 \cdot x} \quad (5.5)$$

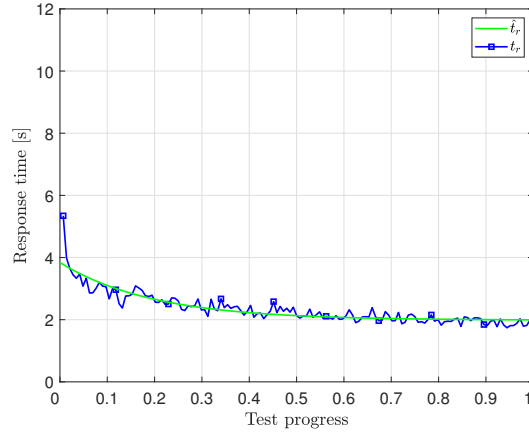


Figure 5.21: Trend of the response time t_r and the estimated response time \hat{t}_r over the test progress. Test progress of 0, resp. 1 indicates the start, resp. end of the test.

where x denotes the test progress ranging the interval $[0; 1]$. The residual sum of squares was 6.32 s^2 , the R^2 value was 82 %, and the root mean squared error was 0.21 s. Hence, the fitted curve approximated well the obtained data. Subjects became familiar with the test procedure and accelerated during the course of the test. Subjects took more time for responding during the 1st series (2.57 s) than during the 2nd series (2.11 s). Based on a Kruskal-Wallis test, there was an effect of the series on the response time (p-value < 0.01). The response times, averaged for each hearing conditions, were 2.17 s (HC0), 2.36 s (HC1), 2.40 s (HC2), 2.53 s (HC3), 2.33 s (HC4), and 2.25 s (HC5). There was a significant effect of the hearing condition on the response time, according to a Kruskal-Wallis test (p-value $\ll 0.001$). By a pairwise Kruskal-Wallis test we identified a significant difference in t_r between hearing condition HC0 and the group HC1 to HC4 (pairwise p-values < 0.001).

5.4.5.2 Replay

Subjects used the replay button for 13.6 % of all stimuli and slightly less often during the 1st series (13.23 %) than during the 2nd series (13.97 %). They relistened male speech (16.07 %) more frequently than white noise (11.13 %). The effect of the sound sample was significant at the 5 % level after a Kruskal-Wallis test. Fitting a 1st order polynomial to the replay rate over the test progress using unweighted linear least squares yielded the function

$$\hat{n}_{\text{rpl}}(x) = 0.00049 \cdot x + 0.1357 \quad (5.6)$$

5.4. LOCALIZATION PERFORMANCE

where $x \in [0; 1]$ indicates the test progress. The resulting residual sum of squares equaled 0.94, R^2 $6.1 \cdot 10^{-6}$, and the root mean squared error 0.057. The low value of R^2 was caused by the high variation of the replay rate along the test. Equation (5.6) shows that the replay rate was constant along the test progress.

5.4.5.3 Preferred sound & Sound imagination

Across all subjects an equal preference of 50 % for both types of stimulus was obtained. Their preferred sound sample corresponded in 72.2 % to the sound they were exposed to during the 2nd series.

Subjects associated the pulsed white noise with sawing wood, stroking a brush or broom across a table or the floor, the sound of an insect, e.g. cicada, the rustle of walking over fallen leaves, and a steam engine.

5.4.5.4 Confusion matrices

The confusion matrices in Table 5.1 are showing the responded quadrant Q_r and the direction of rotation D_r over the tested quadrant Q_t and the direction of rotation D_t . The numerical values and the cell colors are indicating the occurrence rate of the pairs of tested and responded quadrant, resp. direction of rotation. The lower left and upper right 2-by-2 submatrices indicate that there were few ($\leq 3\%$) left-right and mixed-confusions. So, subjects answered either correctly or with front-back confusions exclusively. Independently of the tested quadrant Q_t , subjects perceived sounds in general on the correct side and more often in the back hemisphere (61.75 %) than in the frontal hemisphere (38.75 %). Correct quadrants, resp. directions of rotation were obtained in 56.75 %, resp. 59.00 %, c.f. means of diagonals in Table 5.1. At least two-third of the back sounds ($Q_t \in \{\text{LB}, \text{RB}\}$) were localized in the correct quadrant, while frontal sounds ($Q_t \in \{\text{LF}, \text{RF}\}$) were localized in the correct quadrant in a maximum of 47 %. The confusion matrix of the directions of rotations in Table 5.1b is nearly symmetrical. Counterclockwise (CCW) rotation was perceived correctly in 58 %, clockwise rotation was perceived correctly in 60 %.

5.4. LOCALIZATION PERFORMANCE

		Q_t			
		LF	LB	RB	RF
Q_r	LF	0.44	0.27	0.00	0.01
	LB	0.52	0.70	0.01	0.01
	RB	0.03	0.02	0.66	0.52
	RF	0.02	0.01	0.33	0.47

(a)

		D_t	
		CCW	CW
D_r	CCW	0.58	0.40
	CW	0.42	0.60

(b)

Table 5.1: Confusion table for the quadrants (Table 5.1a) and directions of rotation (Table 5.1b). The values for the quadrants Q_t and Q_r are Left-Front, Left-Back, Right-Back, resp. Right-Front, corresponding to quadrant I to IV of the user interface c.f. Figure 5.20. The values of the directions of rotation D_t and D_r are counterclockwise and ClockWise. The rows correspond to the values of Q_r and D_r responded by the user; the columns correspond to the tested values of Q_t and D_t . The sum of a column might be > 1 due to rounding errors.

5.4.5.5 Localization performance

Subjective sound localization performance was evaluated based on the correct response rate, the left-right, front-back, and mixed confusion rates. Each of the 10368 data sets (288 responses per subject, 36 subjects) were checked if they meet the correct response condition, the left-right, front-back, and mixed confusion condition. If so, it was flagged 1 otherwise 0, respectively for each condition. The correct response condition was satisfied if the responded quadrant Q_r equaled the tested quadrant Q_t and the responded direction of rotation D_r equaled the tested direction of rotation D_t . The front-back, resp. left-right confusion condition was satisfied if the Q_r and Q_t were separated by the frontal, resp. median plane. The mixed confusion condition was satisfied if Q_r and Q_t were diagonally opposing. The rates were obtained by computing the mean value over the flags. Our main interest was the influence of the hearing conditions on the localization performance. Nevertheless, additional factors, such as the sounds, the series, or the groups, were evaluated to gain profound insight about the results.

5.4.5.5.1 Correct responses The correct response rate over all subjects was 51.41 % with a standard deviation of 17.65 %. The correct response rate was 53.59 %, 50.93 %, and 49.71 % in group 0, group 1, and group 2 respectively. A Kruskal-Wallis test showed that the difference between the groups was not significant (p -value > 0.05). When subjects listened to their preferred sound, the correct response rate (51.39 %) was almost the same as when they listened to their non-preferred sound (51.52 %). Subjects

5.4. LOCALIZATION PERFORMANCE

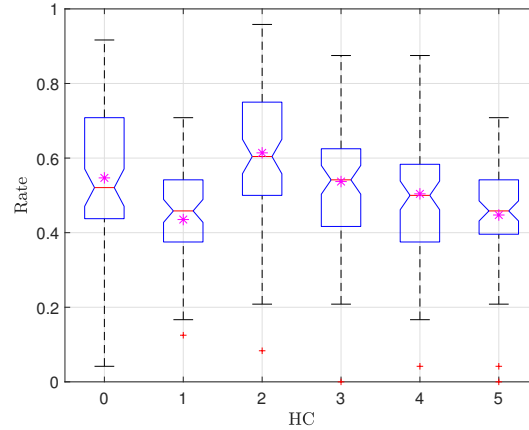


Figure 5.22: Box plot with mean values (asterisks) of correct responses for the hearing conditions HC0 to HC5.

achieved an increased correct response rate with the pulsed white noise (52.76 %) than with the male speech (50.06 %). According to a Kruskal-Wallis test, this difference was not significant (p-value > 0.05). The average correct response rate of 49.73 % during the 1st series was increased to 53.09 % during the 2nd series. The series influenced the correct response rate, accordingly to a Kruskal-Wallis test (p-value < 0.05). The conditional probabilities $P(D_r = D_t | Q_r = Q_t)$ and $P(Q_r = Q_t | D_r = D_t)$ resulted in 0.91 and 0.87.

The distribution of the correct responses rate per hearing conditions is visualized by a box plot in Figure 5.22, including the mean values (c.f. asterisks). Correct responses were obtained in average most often under HC2 with 61.40 %, second most under HC0 with 54.69 %, followed by HC3 (53.70 %), HC4 (50.41 %), and HC5 (44.73 %). Fewest correct responses were obtained under HC1 (43.52 %). A Kruskal-Wallis test showed a significant effect of the hearing conditions on the number of correct responses (p-value $\ll 0.001$). By applying a pairwise test it was identified that HC0 and HC1 led to significant different number of correct responses (p-value < 0.01). Between Prototype A (HC2) and Prototype B (HC2) no difference was obtained (p-value > 0.05). Both significantly increase the number of correct responses with respect to HPD P2 (HC1) (p-values < 0.01) but no difference was obtained with respect to HC0 (p-values $\gg 0.05$). Subjects performed significantly worse with Prototype C (HC4) than with Prototype A (HC2) (p-value < 0.01). Statistically equal numbers of correct responses were obtained between Prototype D (HC5) and Prototype C (HC4), resp. HPD P2 (HC1) (p-values $\gg 0.05$).

5.4. LOCALIZATION PERFORMANCE

The correct response rate over the response time, resp. the replay rate is visualized in Figure 5.23. Both graphs are based on 216 data points, each representing the averaged data of one subject and one hearing condition. PCC for all 216 data equaled in 0.0070, resp. 0.049 not showing any correlation between the correct response rate and the response time, resp. replay rate. These uncorrelated properties can be proven by visual inspection of Figures 5.23a and 5.23b. Regarding the individual hearing conditions, the PCC for the 36 data points of each hearing condition resulted in 0.021, resp. -0.037 (HC0), 0.069, resp. -0.071 (HC1), -0.093 , resp. 0.0079 (HC2), -0.055 , resp. 0.075 (HC3), -0.048 , resp. 0.063 (HC4), and 0.16, resp. 0.026 (HC5) between the correct response rate and the response time, resp. replay rate.

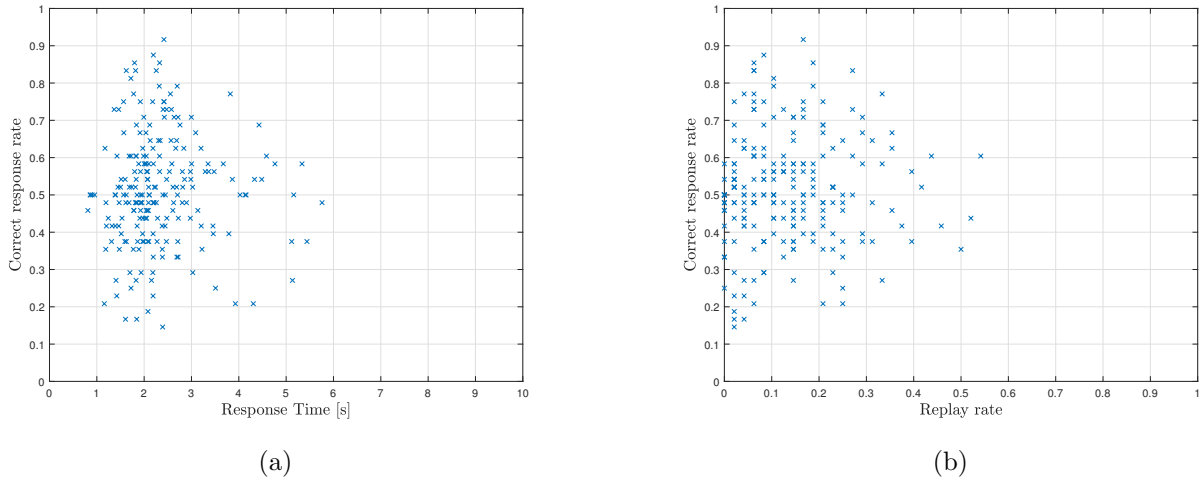


Figure 5.23: Correct response rate over response time (Figure 5.23a), resp. replay rate (Figure 5.23b) for each subject and each hearing condition.

5.4.5.5.2 Front-back confusions We obtained front-back confusions in 40.83 % of all responses. Across the hearing conditions, the averaged front-back confusions ranged the interval $[0.31; 0.49]$, c.f. asterisks in Figure 5.24. The hearing conditions, ordered by increasing front-back confusion rates, are HC2 (30.96 %), HC0 (36.98 %), HC3 (37.32 %), HC4 (41.09 %), HC5 (48.50 %), and HC1 (49.48 %). According to a Kruskal-Wallis test, the hearing conditions significantly affected the front-back confusion rates ($p\text{-value} \ll 0.001$). A pairwise Kruskal-Wallis test on the hearing conditions showed, that we can cluster the hearing conditions in the groups $\{\text{HC0}, \text{HC2}, \text{HC3}\}$, $\{\text{HC4}\}$, and $\{\text{HC1}, \text{HC5}\}$. Hearing conditions of two different groups led to different front-back confusion rates at 1 % level of significance. Subjects erred between frontal and back sound with statistical equal rates under the natural hearing

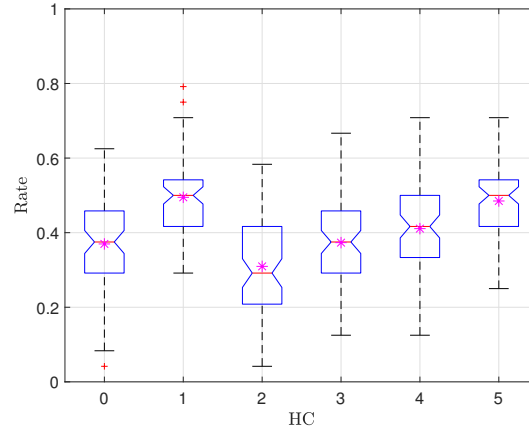


Figure 5.24: Box plot with mean values (asterisks) of front-back confusions for each of the hearing conditions.

(HC0), with Prototype A (HC2), and with Prototype B (HC3). Prototype D (HC5) led to a similar number of front-back confusions as the commercially available headset (HC1). The helmet-based Prototype C (HC4) reduced the front-back confusion rate with respect to the HPD P2 (HC1). Prototype A (HC2) and Prototype B (HC3) resulted in even better performance than Prototype C and further reduced the front-back confusion rate.

Table 5.2 splits the front-back confusion into those where subjects simultaneously responded with inverted directions of rotation ($D_r \neq D_t$), i.e. direction confusion, and where subjects responded with the original directions of rotation ($D_r = D_t$). Front-back confusions with inverted, resp. original directions of rotation were obtained in average in 34.83%, reps. 6.00% of all responses. When subjects responded with a front-back confusion, they inverted in 85.37% the direction of rotation, too. Front-back confusions with inverted directions of rotation strongly depended on the hearing conditions and ranged the interval [0.26; 0.43]. Ordering the hearing conditions by increasing front-back and direction confusions gave HC2, HC0, HC3, HC4, HC5, HC1. The front-back confusion rates with original directions of rotation were approximately constant among the hearing conditions and ranged the interval [0.05; 0.07].

5.4.5.5.3 Left-right confusions Left-right confusions were obtained in 1.47% of all responses. The left-right confusion rates, averaged over all subjects, were 0.58% (HC5), 1.34% (HC1), 1.51% (HC0), 1.62% (HC4), 1.74% (HC2), and 2.02% (HC3), c.f. asterisks in Figure 5.25. Using a Kruskal-Wallis

5.4. LOCALIZATION PERFORMANCE

	HC						mean
	0	1	2	3	4	5	0-6
$D_r \neq D_t$	0.30	0.43	0.26	0.32	0.35	0.43	0.35
$D_r = D_t$	0.07	0.06	0.05	0.06	0.06	0.06	0.06

Table 5.2: The values indicate the front-back confusion rate for each of the six hearing conditions (HC) where subjects responded with the inverted ($D_r \neq D_t$), resp. original ($D_r = D_t$) direction of rotation with respect to the entire number of responses.

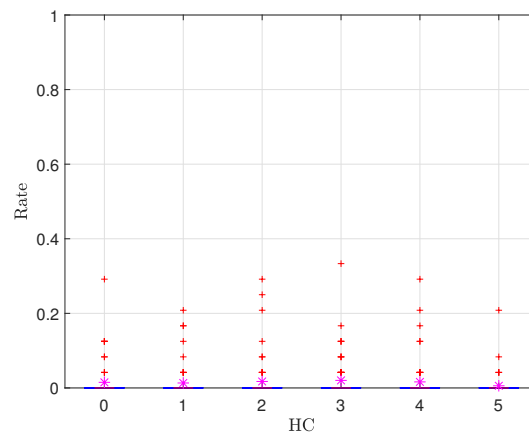


Figure 5.25: Box plot with mean values (asterisks) of left-right confusions for each of the hearing conditions.

test, any statistically significant effects of the hearing conditions on the left-right confusion rate were rejected (p-value of > 0.05). In 88.44 %, resp. 11.56 % of all left-right confusions, D_r was equal to the original, resp. inverted direction of rotation.

5.4.5.5.4 Mixed confusions With 1.14 % of all data, mixed confusions were those errors which we obtained the fewest. The distribution of the mixed confusions are shown in Figure 5.26. Prototype A (HC2) led to the lowest mixed confusion rate of 0.52 %, followed by Prototype D (HC5) with 0.81 %, Prototype C (HC4) with 0.93 %, natural hearing (HC0) with 0.98 %, the commercial headset (HC1) with 1.22 % and Prototype B (HC2) with 2.38 %. A Kruskal-Wallis proved that there was an effect of the hearing conditions on the mixed-confusion rates (p-value < 0.05). Applying a pairwise Kruskal-Wallis test, we found that HC3 resulted in significantly different mixed confusion rates than all other hearing conditions (pairwise p-values < 0.01). In 12.28 % of all mixed confusions, i.e. 0.14 % of all data, the responded direction of rotation equaled the original direction of rotation.

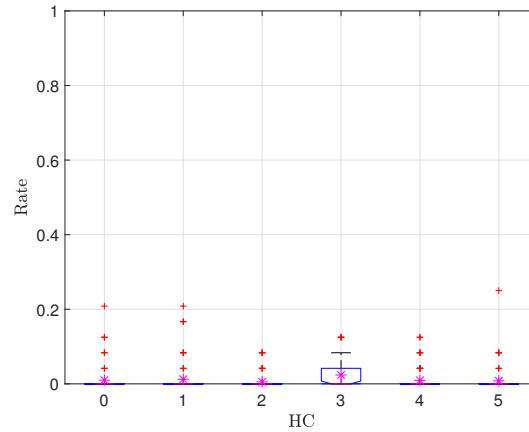


Figure 5.26: Box plot with mean values (asterisks) of mixed confusions for each of the hearing conditions.

5.4.6 Discussion

Subjects aim to finish the 2nd series faster, as they knew the test setup, their task and as the test didn't provide any elements to rise the attention and motivation of the subjects. The acceleration during the series is subjected to be caused by the monotony of the test, fading motivation, and increasing familiarization. Among all listening conditions, HC0 matched closest to natural hearing, causing subjects to respond intuitively and faster than under listening conditions that introduced spectral cues by digital filters.

We expected the replay button to be used with varying frequency along the course of the test, e.g. increasing frequency due to fatigue or decreasing frequency due to adaption. In contrary we obtained that subjects relisted sounds with constant rate along the test progress. On the one hand, it seems that fatigue and adaption balanced each other out. On the other hand, subjects did not want to prolong the time spend on the test by using the replay button. Neither the frequency with which the replay button was used nor thinking about the sound direction seemed to affect the localization performance. This rather confirms that subjects followed the instruction to choose their answer instinctively and to use the replay button only in exceptional cases. In contrast, the dual task of recognizing the quadrant and the direction of rotation may have been difficult for subjects and may have required high cognitive effort, leading to faster fatigue.

Once subjects in group 2 received the notification, they seem to discard their judgments about

5.4. LOCALIZATION PERFORMANCE

the spectral cues and their decision rule for front-back discrimination and aim to identify a new rule for the upcoming subset. Subjects of group 1 seem to notice differences in the spectrum and aim to identify a new decision rule. In the case of group 0 subjects seem to keep the same decision rule along the entire test, leading to an increased correct response rate.

Both types of stimuli are repetitive, which might be more severe for male speech as for pulsed white noise. In contrast, subjects might be more familiar with speech which might be perceived more pleasant than white noise. The male speech sample is unambiguous, while the pulsed noise sample left room for imagination about the sound source. The possible drawback of annoyance of the noise sample was compensated by the subjective imagination. The constant peak amplitude of the wide band pulsed noise slightly increases the rate of correct responses compared to the male speech, which has a temporally decreasing peak amplitude and a limited bandwidth.

Subjects didn't know much about what they were faced to before performing the 1st series. In contrast, for the 2nd series, they had gained some experience and knew about the proceeding of the test, which made them more self-confident and led to an increased correct response rate. Some subjects seemed to think about which sound sample they prefer, but the majority simply chose the last sound heard as their preferred sound.

Due to the test design, the responded quadrant and the responded direction of rotation are related and if one is correct, it is very likely that the other is correct, too. Correct responses and front-back confusions account for 92.24% of all responses and the PCC between them is -0.99 . Thus, subjects responded either correctly or with a front-back confusion. Sound localization performance is therefore primarily affected by the front-back confusions associated with direction confusions. Mirroring the perceived quadrant at the frontal plane while preserving the perception of a movement from lateral to median positions or vice versa, results inevitably in an inversion of the direction of rotation ($D_r \neq D_t$). Subjects who responded with a front-back confusion but with the original direction of rotation ($D_r = D_t$) perceived the stimuli in the opposite direction to the actual lateral-to-median or median-to-lateral movement. These issues are caused by corruptions of the provided ILD and ITD.

When subjects listened with the natural hearing (HC0) they listened with the natural HRTF of the dummy head rather than their individual HRTF. Hence, the hearing condition HC0 corresponds to sound localization with a non-individualized HRTF. Prototype C (HC4), that used the HRTF of dummy head KEMAR, and Prototype D (HC5), that introduced the spectral cues of the outer ear-

like cavity, were also introducing non-individual HRTFs. It seems that subjects could not interpret the spectral cues of the non-individual HRTFs as directional information. The diffraction around the helmet additionally blurs the position of the sound source (Prototype C). The symmetry of the spectral cues and the lack of spectral front-back difference with Prototype D are further increasing the front-back confusion rates. In contrast, the microphone distance of Prototype D introduced in larger ITD values than the natural outer ear distance, resulting in more lateral localization of sounds. The simple signal processing with directional microphones and a HSF seems to be most favorable over all prototypes. Even though the spectral cues were not very detailed, subjects managed to match a direction to the provided acoustic information.

Prototype B (HC3) struggles of the error-proneness of the ambisonic technology on the position and orientation of the recording microphone. Due to the non-ideal recording, the contralateral virtual loudspeakers received strong signals than required, which shifted the stimulus. Further, the deployed technique for mixing the four virtual loudspeakers signals to one stereo signal provides only weak ILD, but no ITD. In the case of discordance, the ITD dominates over the ILD [205]. This led to stimulus being perceived at median positions and making subjects randomly choosing between a left and right quadrant. Advanced ambisonic decoders, rectangular or hexagonal virtual loudspeaker setups, and HRTFs might enhance sound localization [206, 207]. This in turn increases the computational power. Further, the choice of the HRTF must be well made to match the subjects.

The listening test simulated dynamic localization cues by a relative source-head movement. Therefore, localization performance should have increased compared to the localization test in Section 2.3 with its static cues and source-head positioning. Contrary to our expectations, we obtained with HPD P2 less front-back confusions with static (37%) than with dynamic (49%) cues. The listening test didn't provide vestibular cues and experience of daily life made humans localize sounds in the back, if there is no object in the field of view which they can identify as sound source. To overcome this psychological effect, the prototypes need to be upgraded to fully functional HPDs and their evaluation should be conducted as a "live" localization test in the semi-anechoic chamber, like the localization test in Section 2.3. By this, subjects listen with their individual HRTF, the HC0 hearing condition would actually correspond to natural listening, and the drawback of dummy head recordings would be eliminated [192].

5.5 Conclusion

Four different designs of earmuff HPDs were presented in the beginning of this chapter. They are based on cardioid microphones which were mounted on the shells (Prototype A) and on a ballistic helmet (Prototype C), equipped with an ambisonic microphone (Prototype B), and a concha-like cavity in the outside of the shells (Prototype D). Except Prototype D whose cavities filter the incoming sound similar to the human outer ear, the prototypes introduce spectral cues by filtering digitally the microphone signals. The localization performance of the prototypes was evaluated by a subjective listening test and compared to the localization performance with the HRTF of a dummy head and a commercially available headset HPD P2, c.f. Figure 2.2.

Even though the spectral cues, which are introduced by the prototypes, slightly differ to each other, they show common characteristics, which are corresponding to Blauert's boosted band between 1.86 kHz and 7.03 kHz [40]. The spectral differences between frontal and back sound incidence were clearly noticeable by applying frequency analysis. However, it turned out that for Prototype D the combination of the cavity and the microphone positioning led to similar spectral cues for frontal and back sound incidence. The spectral difference should be enhanced by studying more asymmetric designs, e.g. geometries with less material at the front of the shell.

The listening test supported the findings from the frequency analysis. Subjects localized sounds the best with Prototype A, followed by Prototype B. They both boost high frequencies of frontal sounds. These results are in coherence with Frank et al. and Brungart et al. [76, 181]. Contrarily to Rubak et al. who reported enhanced sound localization with cavities in the shells, we observed reduced sound localization with Prototype D [132]. The reduced spectral differences between frontal and back sound incidence are identified to cause the reduced performance. The Prototype C is a promising good approach but requires amelioration of the microphone array to provide more unidirectional recordings.

Even though Prototype D does not introduce spectral front-back differences while HPD P2 does introduce such differences, both devices led to similar localization performance. Hence, to allow proper sound localization it is not sufficient to simply provide any spectral cues. It is important to provide spectral cues that users can correctly interpret and assign directions to. Here we are in accordance with Brungart et al. who demand from HPD not only providing a wide system bandwidth to provide enough spectral cues but also proper spectral information [181].

5.5. CONCLUSION

The design of the listening test was subjected to influence the obtained localization performance. The sound source was turning around the subjects, but the relative position to the head was fix. When subjects turned their heads, the sound source rotated in the same way. Hence, the vestibular cues conflict with the constant head-to-sound positioning. Further, subjects realized that their visual environment, i.e. the office pod, and the acoustic environment were decoupled. Hence, sound sources are hardly perceived in the front. To overcome these drawbacks, the prototypes need to be designed and realized as fully functional HPDs which must be reevaluated with real sound sources and a VAE such as used in Section 1.4.2.2.

Concerning the usability of the presented devices, Prototype B is the least practical in field. In its next stage of development, it should be envisaged to mount the ambisonic microphone the closest possible to the headband. Another approach would be to mount two ambisonic microphones on the top of the shells, one microphone on each shell, just next to the head, slightly above the outer ear. This requires evaluation of the effect of the head on the acoustic field in such proximity which might be promisingly compensated by binaural techniques. Neither Prototype A nor Prototype C are yet fully developed. Avoiding protruding microphones on Prototype A could be achieved by redesigning the shells with appropriate cavities. Flush mounting the microphone on the ballistic helmet of Prototype C retains any directional characteristic of the microphones. Therefore, it seems more promising to mount a dense array of miniature digital microphones with tiny dimensions on the ballistic helmet. Optimized beam forming algorithms seem to be suitable to scan the surrounding environment for sound sources, to focus on them separately, and to restore them with proper directional cues under the hearing protection. Limited computational power and energy supply may restrict these algorithms on either a limited number of simultaneously processed sound sources or low angular resolution. This in turn conflicts with 360° surveillance of a large number of simultaneous sources. Prototype D seems to be the best design in terms of usability as there aren't any parts which are sticking out. The design of the shells requires optimization to allow compatibility with ballistic helmets.

Conclusion

The overall aim of this work was to propose principles and solutions for acoustically “transparent” hearing protections. The auditory environment was intended to be reconstructed under these hearing protections so that the user not only perceives the sounds acoustically, but also hears where the sounds are located. For this we wanted to gain insight on, first, the HRTF modifications introduced by HPDs and the resulting localization performance and, second, about possible design approaches. Third, on methods for capturing the directional cues of environmental sounds and, forth and finally, on techniques for reconstructing the sounds with directional cues under the HPD.

6.1 Results

An existing loudspeaker based VAE was completely redesigned, including the installation of 16 professional loudspeakers and their driver electronics, and developing a system for head motion tracking. A dedicated software was implemented, which provides independent control on all loudspeaker signals, communicates with the head motion tracker, and conducts listening tests in the VAE.

Measurements of HPD induced modifications in the HRTFs of three different dummy heads showed that the modifications highly depend on the deployed HPD. Earplugs showed little effect on the HRTFs, while earmuffs strongly changed the HRTFs along the entire frequency range. The results of a subjective sound localization test, which was conducted in the loudspeaker based VAE with 40 subjects, each testing 4 HPDs, are in line with the results of the HRTF measurements. An inverse proportional relation between the HRTF modifications and the localization performance was identified, i.e. the more the HRTF is modified the worse subjects localize sound sources. We obtained that the tested earplugs, resp. active HPDs were preferable over earmuffs, resp. passive HPDs. Numerical simulations, based on the 3D scans of the dummy heads, have shown that the outer ear exclusively

introduces the individual spectral cues which are required for front-back discrimination. Hence, to maintain localization performance with earplugs, they should cover the least possible the outer ear. This can be achieved using miniaturization technology, already used for hearing aids, to construct tiny earplugs that are placed in the ear canal. Since this approach is not applicable for earmuffs, additional solutions were presented and evaluated.

An analytic model was developed which describes the spectral front-back differences of measured and simulated HRTFs. The model is controlled only by the parameter ϕ , i.e. the azimuth angle, and returns the filter coefficients for a 10th order IIR filter. The model is based on five 2nd order IIR peak filters. They are combined to optimally approximate the angle-dependent spectral difference of the underlying HRTFs between pairwise positions symmetric about the interaural axis. The spectral error between the model and the underlying HRTFs is inferior to 2 dB. Two series of listening tests with 30 (1st series) and 33 subjects (2nd series) were conducted to assess the localization performance of sounds in a headphone based VAE. Digital filters introduced directional front-back information to the stimuli. A large variety of filters was tested including an High-Shelf Filter (HSF), the developed model for spectral front-back difference, peaking filters, and HRTFs. The design of the listening test, i.e. the unforced choice design, was identified to bias the localization performance. Further, the similarity of the filters which were deployed in the 1st series led to similar localization performance among the filters. The 2nd series showed that the low order (max. 10th order) IIR filter led to better front-back differentiation than detailed high order non-individual HRTFs.

Four designs of advanced earmuff HPDs were developed, and the corresponding prototypes were set up. In general, the prototypes recorded the surrounding sound field with directional microphones. They reconstruct the sound field with directional information under the earmuff, by applying a subset of the previously evaluated numerical filters. Prototype A consists of an array of six microphones mounted on the exterior of the earmuffs, Prototype B consists of a 1st order ambisonic microphone which is centered over the head, and Prototype C consists of a microphone array attached to a ballistic helmet. Prototype D comprises a concha like cavity, which introduces the spectral cues like the human outer ears without the need of numerical filters. Measurements showed that Prototype D introduces least spectral front-back cues. Further, Prototype A to C introduce spectral front-back differences which are most distinct between 2 kHz and 8 kHz. These characteristics are similar to natural HRTFs and the positions of Blauert's directional bands. The sound localization performance with these

prototypes was evaluated in a listening test with 36 participants. With respect to a commercially available active earmuff, all prototypes reduced the number of front-back errors and increase sound localization performance. Compared to a non-individual HRTF, only Prototype A led to a reduced number of front-back errors.

The response to the initial problem of “transparent” hearing protections can be summarized as: Miniaturized, active earplugs are required to allow the user to perceive the environmental sounds correctly positioned in space and enhance situational awareness. With active earmuffs it is possible to restore the directional information by applying simple digital filters to the recorded sound field signals. Additionally, audio which is transmitted through radio communication can be enriched with directional cues by making use of the technique of directional information reconstruction of earmuffs.

6.2 Perspectives

The HPD prototypes and the correspondingly obtained subjective localization performance requires validation. The sensor-based prototypes need to be upgraded to fully functional HPDs by adding signal processing electronics and internal reproduction components and reviewing the acoustic attenuation. A real-time subjective localization test, similar to the initial localization test, has to be followed in a loudspeaker based VAE. Prior to the subjective localization test it is worth predicting the localization performance under these HPDs with the models developed by Joubaud [166]. Thus, the localization performance and HPD rankings we observed can be verified by two independent methods. Future designs of earmuffs may also be tested, which consists of a multichannel reproduction system, similar to Pomberger et al. [208]. This approach seems to be the most promising solution for augmenting the localization performance with HPDs as it incorporates the users’ individual HRTF. Such prototypes should be manufactured and evaluated for verification. Earplugs will benefit of miniaturization technologies. By reducing their geometries, they can be inserted into the ear canal without covering the outer ear. Consequently, they do not modify the user’s individual spectral cues. They have to block the ear canal in order to provide acoustic protection and might be designed as either rigid, custom molded devices, or multiuser devices. Earmuffs might also benefit from miniaturized microphones, as they reduce the weight of the protection and the impact on the sound field.

Involving human subjects in further studies might carry out a much more profound knowledge

about the interaction between HPDs, HRTFs, and sound localization performance. Measuring the spectral modifications caused by HPDs in the individual HRTFs of subjects allows to evaluate the variance of the spectral modifications between individuals. Additionally, the subjective, individual sound localization performance can be confronted with the HPD induced modifications in the individual HRTFs of the subjects rather than with the HRTFs of dummy heads. Further, human subjects are necessary for upcoming work on the approximation of individual HRTFs by combining a generic transfer function, e.g. of an ellipsoid or head without ears, with the individual PRTF. First, investigation must be done on the relation between individual HRTF, individual PRTF, and individual anatomy. Second, algorithms need to be developed and enhanced for combining a generic transfer functions with an individual PRTF in order to approximate the original, individual HRTF. Third and finally, the combined HRTFs need to be evaluated in terms of sound localization performance and compared to the original, individual HRTF. In combination with a 3D scanner and simulation software this technique would allow to develop a portable field system for obtaining individual HRTFs for a large public. Complex HRTF measurements in anechoic chamber would get redundant and personnel devices such as hearing protections, communications systems, or headphones can thus be adapted anywhere, independent of laboratory facilities.

The design of Prototype D is promising but requires additional revision. To overcome the issue of limited localization performance, the earmuff with the concha like cavity requires updates on its geometry. New propositions on the geometry should be examined, such that asymmetric spectral cues are generated. In particular, the frontal part of the shell is considered to provide potential for optimization. The geometry of the cavity can be further updated considering the advanced pinna simulator from [147] or the individual anatomy of the outer ear of human subjects. The generalized geometry from [147] results in multi-user, non-individual prototypes. the application of the individual anatomy leads to custom, i.e. individualized, prototypes that resolve the issue with non-individual HRTF. This is extreme costly as the individual geometries of the outer ears have to be scanned and prototypes for each subject need to be manufactured. To reduce these costs, a system can be imaged which is based on an earmuff with a detachable cavity, and users mount their own, individualized cavity to the earmuff.

Bibliography

- [1] G. J. Augustine, D. M. Chikaraishi, M. D. Ehlers, G. Einstein, D. Fitzpatrick, W. C. Hall, E. Jarvis, L. C. Katz, J. Kauer, A.-S. LaMantia, J. O. McNamara, R. D. Mooney, M. A. L. Nicolelis, D. Purves, P. H. Reinhart, S. A. Simon, J. H. P. Skene, J. Voyvodic, L. E. White, and S. M. Williams, *Neuroscience*. Sinauer Associates, Inc., third ed., 2004.
- [2] A. Thompson and B. N. Taylor, *Guide for the Use of the International System of Units (SI)*. NIST Special Publication 811, 2008 ed., Nov. 2008.
- [3] D. Havelock, S. Kuwano, and M. Vorländer, eds., *Handbook of Signal Processing in Acoustics*, vol. 1. Springer, 2008.
- [4] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*. Springer-Verlag, 1990.
- [5] F. Paulsen and J. Waschke, *Sobotta Atlas der Anatomie des Menschen: Kopf, Hals und Neuroanatomie*, vol. 3. Elsevier Health Sciences, fifteenth ed., 2011.
- [6] M. Schünke, E. Schulte, U. Schumacher, M. Voll, and K. Wesker, *Prometheus – Lernatlas der Anatomie: Kopf, Hals und Neuroanatomie*, vol. 4, ch. Organe und ihre Leitungsbahnen, pp. 142–157. Georg Thieme Verlag, 2015. In German.
- [7] J. Waschke, T. M. Böckers, F. Paulsen, S. Winkler, K. Dalkowski, J. Mair, and S. Klebe, *Sobotta Lehrbuch der Anatomie*, vol. 2. Elsevier Health Sciences, second ed., 2019.
- [8] P. L. M. Johannesma, “The pre-response stimulus ensemble of neurons in the cochlear nucleus,” in *Symposium on Hearing Theory, 1972*, IPO, 1972.

BIBLIOGRAPHY

- [9] R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, “Complex sounds and auditory images,” in *Auditory Physiology and Perception* (Y. Cazals, K. Horner, and L. Demany, eds.), pp. 429–446, Pergamon, 1992.
- [10] H. F. Olson, *Music, physics and engineering*, vol. 1769. Courier Corporation, 1967.
- [11] E. Zwicker, G. Flottorp, and S. S. Stevens, “Critical band width in loudness summation,” *The Journal of the Acoustical Society of America*, vol. 29, pp. 548–557, May 1957.
- [12] E. Zwicker, “Subdivision of the audible frequency range into critical bands (Frequenzgruppen),” *The Journal of the Acoustical Society of America*, vol. 33, p. 248, Feb. 1961.
- [13] International Organization for Standardization, “ISO 226:2003 – Acoustics – Normal equal-loudness-level contours,” Sept. 2003.
- [14] K. A. B. Knobel and T. G. Sanchez, “Loudness discomfort level in normal hearing individuals,” *Pro-fono: revista de atualizacao cientifica*, vol. 18, no. 1, pp. 31–40, 2006.
- [15] R. A. Bentler and L. J. Cooley, “An examination of several characteristics that affect the prediction of OSPL90 in hearing aids,” *Ear and Hearing*, vol. 22, no. 1, pp. 58–64, 2001.
- [16] U. Katzenell and S. Segal, “Hyperacusis: review and clinical guidelines,” *Otology & neurotology*, vol. 22, no. 3, pp. 321–327, 2001.
- [17] B. C. J. Moore, *An introduction to the psychology of hearing*. Emerald Group Publishing, sixth ed., 2012.
- [18] S. A. Gelfand, *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2016.
- [19] H. Fletcher and W. A. Munson, “Loudness, its definition, measurement and calculation,” *The Journal of the Acoustical Society of America*, vol. 5, pp. 82–108, Oct. 1933.
- [20] D. W. Robinson and R. S. Dadson, “A re-determination of the equal-loudness relations for pure tones,” *British Journal of Applied Physics*, vol. 7, no. 5, p. 166, 1956.
- [21] R. M. Aarts, “A comparison of some loudness measures for loudspeaker listening tests,” *Journal of the Audio Engineering Society*, vol. 40, no. 3, pp. 142–146, 1992.

BIBLIOGRAPHY

- [22] International Organization for Standardization, “ISO 389-7:2019 – Acoustics – Reference zero for the calibration of audiometric equipment – part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions,” Oct. 2019.
- [23] International Electrotechnical Commission, “61672-1: 2013 Electroacoustics – Sound level meters – Part 1: Specifications,” 2013.
- [24] American National Standards Institute, “ANSI S1.4:2001 Specification for sound level meters.” Acoustical Society of America.
- [25] American National Standards Institute, “ANSI S1.42:2001 Design response of weighting networks for acoustical measurement.” Acoustical Society of America.
- [26] Berufsgenossenschaft Holz und Metall, “DGUV Information 209–023: Lärm am Arbeitsplatz,” Nov. 2013. In German.
- [27] I. Pollack and J. M. Pickett, “Cocktail party effect,” *The Journal of the Acoustical Society of America*, vol. 29, no. 11, pp. 1262–1262, 1957.
- [28] B. Arons, “A review of the cocktail party effect,” *Journal of the American Voice I/O Society*, vol. 12, no. 7, pp. 35–50, 1992.
- [29] John William Strutt, Baron Rayleigh, *The theory of sound*. MacMillan and Co. London, 1877.
- [30] C. L. Searle, L. D. Braida, D. R. Cuddy, and M. F. Davis, “Binaural pinna disparity: another auditory localization cue,” *The Journal of the Acoustical Society of America*, vol. 57, no. 2, pp. 448–455, 1975.
- [31] Wolfram Alpha LLC, “Wolfram|Alpha.” <https://www.wolframalpha.com/input/?i=diameter+human+head>. (retrieved May, 22 2018).
- [32] J. C. Middlebrooks, “Chapter 6 - Sound localization,” in *The Human Auditory System* (M. J. Aminoff, F. Boller, and D. F. Swaab, eds.), vol. 129 of *Handbook of Clinical Neurology*, pp. 99–116, Elsevier, 2015.
- [33] E. Macpherson and J. Middlebrooks, “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,” *The Journal of the Acoustical Society of America*, vol. 111, pp. 2219–2236, June 2002.

BIBLIOGRAPHY

- [34] J. C. Middlebrooks and D. M. Green, “Sound localization by human listeners,” *Annual review of psychology*, vol. 42, no. 1, pp. 135–159, 1991.
- [35] J. Zwislocki and R. S. Feldman, “Just noticeable differences in dichotic phase,” *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 860–864, 1956.
- [36] R. G. Klumpp and H. R. Eady, “Some measurements of interaural time difference thresholds,” *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 859–860, 1956.
- [37] A. W. Mills, “On the minimum audible angle,” *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.
- [38] J. Blauert, *The technology of binaural listening*. Springer, 2013.
- [39] A. M. Colman, *A dictionary of psychology*. Oxford University Press, fourth ed., 2015.
- [40] J. Blauert, “Sound localization in the median plane,” *Acoustica*, vol. 22, pp. 205–213, 1969.
- [41] E. H. A. Langendijk and A. W. Bronkhorst, “Contribution of spectral cues to human sound localization,” *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.
- [42] P. H. Myers, “Three-dimensional auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization,” 1989. US Patent 4817149.
- [43] J. Hebrank and D. Wright, “Spectral cues used in the localization of sound sources on the median plane,” *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.
- [44] C.-J. Tan and W.-S. Gan, “User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems,” *Electronics letters*, vol. 34, no. 25, pp. 2387–2389, 1998.
- [45] K. H. Ngan, “Optimizing front/back confusion rates in sound localisation performance: cluster analyses and experimental studies,” Master’s thesis, Hong Kong University of Science and Technology, June 2005.
- [46] S. Ghorbal, T. Auclair, C. Soladié, and R. Segurier, “Pinna morphological parameters influencing HRTF sets,” in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, 2017.

- [47] F. L. Wightman and D. J. Kistler, “Headphone simulation of free-field listening. I: stimulus synthesis,” *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 858–867, 1989.
- [48] A. Abaza and A. Ross, “Towards understanding the symmetry of human ears: A biometric perspective,” in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–7, IEEE, 2010.
- [49] J. Fels, “How do head-related transfer functions of children depend on growth,” in *Proceedings of the Joint Congress CFA/DAGA '04*, pp. 937–938, 2004.
- [50] V. Zimpfer and D. Sarafian, “Impact of hearing protection devices on sound localization performance,” *Frontiers in neuroscience*, vol. 8, no. 135, pp. 167–176, 2014.
- [51] P. M. Hofman, J. G. A. van Riswick, and A. J. van Opstal, “Relearning sound localization with new ears,” *Nature neuroscience*, vol. 1, no. 5, pp. 417–421, 1998.
- [52] D. W. Batteau, “The role of the pinna in human localization,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 168, no. 1011, pp. 158–180, 1967.
- [53] J. Moreira, L. Gros, R. Nicol, I. Viaud-Delmon, C. Le Prado, and S. Natkin, “Binaural sound rendering improves immersion in a daily usage of a smartphone video game,” in *EAA Spatial Audio Signal Processing Symposium*, pp. 79–84, Sept. 2019.
- [54] M. Geronazzo, A. Rosenkvist, C. Eriksen, M.-H. Kirstine, J. Køhlert, M. Valimaa, M. Vittrup, and S. Serafin, “Creating an audio story with interactive binaural rendering in virtual reality,” *Wireless Communications and Mobile Computing*, pp. 1–14, July 2019.
- [55] D. Schönstein and B. F. G. Katz, “HRTF selection for binaural synthesis from a database using morphological parameters,” in *International Congress on Acoustics (ICA)*, pp. 1–6, Aug. 2010.
- [56] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.
- [57] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, “Binaural technique: Do we need individual recordings?,” *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

BIBLIOGRAPHY

- [58] A. Blum, B. F. G. Katz, and O. Warusfel, “Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training,” in *Proceedings of Joint Meeting of the German and the French Acoustical Societies (CFA/DAGA '04)*, Strasbourg, France, pp. 1225–1226, 2004.
- [59] G. Parseihian and B. F. G. Katz, “Rapid head-related transfer function adaptation using a virtual auditory environment,” *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2948–2957, 2012.
- [60] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Rendering localized spatial audio in a virtual auditory space,” *IEEE Transactions on multimedia*, vol. 6, no. 4, pp. 553–564, 2004.
- [61] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The cipic hrtf database,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 99–102, IEEE, 2001.
- [62] W. G. Gardner and K. D. Martin, “HRTF measurement of a KEMAR,” *The Journal of the Acoustical Society of America*, vol. 67, pp. 3907–3908, June 1995.
- [63] H. Braren and J. Fels, “A high-resolution individual 3D adult head and torso model for HRTF simulation and validation: HRTF measurement,” tech. rep., Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Germany, 2020.
- [64] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, “Fast head-related transfer function measurement via reciprocity,” *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2202–2215, 2006.
- [65] N. A. Gumerov, A. E. O’Donovan, R. Duraiswami, and D. N. Zotkin, “Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation,” *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 370–386, 2010.
- [66] H. Ziegelwanger, P. Majdak, and W. Kreuzer, “Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization,” *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 208–222, 2015.

- [67] A. Meshram, R. Mehra, and D. Manocha, “Efficient HRTF computation using adaptive rectangular decomposition,” in *AES 55th International Conference , Helsinki, Finland*, pp. 1–9, 2014.
- [68] B. U. Seeber and H. Fastl, “Subjective selection of non-individual head-related transfer functions,” in *International Conference on Auditory Display*, Georgia Institute of Technology, 2003.
- [69] E. A. Torres-Gallegos, F. Orduna-Bustamante, and F. Arámbula-Cosío, “Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database,” *Applied Acoustics*, vol. 97, pp. 84–95, 2015.
- [70] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, “HRTF personalization using anthropometric measurements,” in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*, pp. 157–160, IEEE, 2003.
- [71] A. Kuhn, M. Rothbucher, and K. Diepold, “HRTF customization by regression,” tech. rep., Technische Universität München, Lehrstuhl für Datenverarbeitung, Apr. 2014.
- [72] P. Paukner, M. Rothbucher, and K. Diepold, “Sound localization performance comparison of different HRTF-individualization methods,” tech. rep., Technische Universität München, Lehrstuhl für Datenverarbeitung, Apr. 2014.
- [73] W. L. Martens, “Principal components analysis and resynthesis of spectral cues to perceived direction,” in *Proc. of the International Computer Music Conference, Champaigne-Urbana, IL*, pp. 1–8, 1987.
- [74] D. J. Kistler and F. L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.
- [75] P. Nowak, V. Zimpfer, and U. Zölzer, “Automatic approximation of head-related transfer functions using parametric IIR filters,” in *DAGA 2020 - 46. Jahrestagung für Akustik, Hannover, Germany*, pp. 213–216, 2020.
- [76] M. Frank and F. Zotter, “Simple reduction of front-back confusion in static binaural recording,” in *DAGA 2018 - 44. Jahrestagung für Akustik, Munich, Germany*, pp. 329–332, 2018.

BIBLIOGRAPHY

- [77] J. Blauert, “Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene,” *Acoustica*, vol. 20, pp. 200–206, 1968.
- [78] T. N. Roth, “Chapter 20 - Aging of the auditory system,” in *The Human Auditory System* (M. J. Aminoff, F. Boller, and D. F. Swaab, eds.), vol. 129 of *Handbook of Clinical Neurology*, pp. 357–373, Elsevier, 2015.
- [79] Y. H. Park, S.-H. Shin, S. W. Byun, and J. Y. Kim, “Age-and gender-related mean hearing threshold in a highly-screened population: The korean national health and nutrition examination survey 2010 – 2012,” *PLoS One*, vol. 11, no. 3, pp. 1–13, 2016.
- [80] K. Kurakata, T. Mizunami, H. Sato, and Y. Inukai, “Effect of ageing on hearing thresholds in the low frequency region,” *Journal of low frequency noise, vibration and active control*, vol. 27, no. 3, pp. 175–184, 2008.
- [81] I. Maccà, M. L. Scapellato, M. Carrieri, S. Maso, A. Trevisan, and G. B. Bartolucci, “High-frequency hearing thresholds: effects of age, occupational ultrasound and noise exposure,” *International archives of occupational and environmental health*, vol. 88, no. 2, pp. 197–211, 2015.
- [82] T. L. Wiley, R. Chappell, L. Carmichael, D. M. Nondahl, and K. J. Cruickshanks, “Changes in hearing thresholds over 10 years in older adults,” *Journal of the American Academy of Audiology*, vol. 19, no. 4, pp. 281–292, 2008.
- [83] M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld, “Auditory and non-auditory effects of noise on health,” *The lancet*, vol. 383, no. 9925, pp. 1325–1332, 2014.
- [84] M. Spreng, “Possible health effects of noise induced cortisol increase,” *Noise Health [serial online]*, vol. 2, no. 7, pp. 59–63, 2000. Available from: <http://www.noiseandhealth.org/text.asp?2000/2/7/59/31741>.
- [85] K. Hopkins, “Chapter 27 - Deafness in cochlear and auditory nerve disorders,” in *The Human Auditory System* (M. J. Aminoff, F. Boller, and D. F. Swaab, eds.), vol. 129 of *Handbook of Clinical Neurology*, pp. 479–494, Elsevier, 2015.
- [86] L. E. Humes, L. M. Joellenbeck, J. S. Durch, *et al.*, *Noise and military service: Implications for hearing loss and tinnitus*. National Academies Press, 2005.

BIBLIOGRAPHY

- [87] M. Sliwinska-Kowalska, A. Davis, *et al.*, “Noise-induced hearing loss,” *Noise and Health*, vol. 14, no. 61, p. 274, 2012.
- [88] Arbeitskreis 1.6 “Lärm” des Ausschusses Arbeitsmedizin der Gesetzlichen Unfallversicherung, “Handlungsanleitung für die arbeitsmedizinische Vorsorge (BGI/GUV-I 504-20).” www.dguv.de/publikationen, Nov. 2011. In German.
- [89] G. Smedje, M. Lundén, L. Gärtner, H. Lundgren, T. Lindgren, *et al.*, “Hearing status among aircraft maintenance personnel in a commercial airline company,” *Noise and Health*, vol. 13, no. 54, p. 364, 2011.
- [90] European Parliament and Council, “Directive 2003/10/EC of the European Parliament and of the council,” *Official journal of the European Union*, vol. 46, pp. 38–44, Feb. 2003. Accessed May 19, 2020.
- [91] The German Federal Ministry of Justice, “Verordnung zum Schutz der Beschäftigten vor Gefährdungen durch Lärm und Vibrationen (Lärm- und Vibrations-Arbeitsschutzverordnung – LärmVibrationsArbSchV),” *German legal text*, Mar. 2007. In German.
- [92] Arbeitsring Lärm der Deutsche Gesellschaft für Akustik e.V., “Geräuschkulissen.” <https://www.ald-laerm.de/publikationen/hoerbeispiele/geraeuschkulissen/>. Accessed May 19, 2020.
- [93] World Health Organization, “Hearing loss due to recreational exposure to loud sounds: a review,” tech. rep., World Health Organization, 2015.
- [94] E. H. Berger and J. G. Casali, *Hearing Protection Devices*, vol. 2, ch. 81, pp. 967–981. John Wiley & Sons, Incorporated, 2007.
- [95] E. H. Berger, R. W. Kieper, and D. Gauger, “Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways,” *The Journal of the Acoustical Society of America*, vol. 114, no. 4, pp. 1955–1967, 2003.
- [96] C. J. Fackler, E. H. Berger, W. J. Murphy, and M. E. Stergar, “Spectral analysis of hearing protector impulsive insertion loss,” *International journal of audiology*, vol. 56, no. sup1, pp. 13–21, 2017.

BIBLIOGRAPHY

- [97] R. T. Sataloff and J. Sataloff, eds., *Occupational Hearing Loss*. CRC Press Taylor & Francis Group, 2006.
- [98] W. J. Murphy, D. C. Byrne, and J. R. Franks, “Firearms and hearing protection,” *Hearing Review*, vol. 14, no. 3, p. 36, 2007.
- [99] S. M. Abel and P. Odell, “Sound attenuation from earmuffs and earplugs in combination: maximum benefits vs. missed information,” *Aviation, space, and environmental medicine*, vol. 77, no. 9, pp. 899–904, 2006.
- [100] A. Behar and H. Kunov, “Insertion loss from using double protection,” *Applied acoustics*, vol. 57, no. 4, pp. 375–385, 1999.
- [101] S. M. Abel and N. M. Armstrong, “The combined sound attenuation of earplugs and earmuffs,” *Applied Acoustics*, vol. 36, no. 1, pp. 19–30, 1992.
- [102] E. H. Berger, “Laboratory attenuation of earmuffs and earplugs both singly and in combination,” *American Industrial Hygiene Association Journal*, vol. 44, no. 5, pp. 321–329, 1983.
- [103] C. Tubb, S. Mercy, and S. James, “Investigating double hearing protection using human subjects,” *New Directions for Improving Audio Effectiveness*, pp. 1–12, 2005.
- [104] M. Toivonen, R. Pääkkönen, S. Savolainen, and K. Lehtomäki, “Noise attenuation and proper insertion of earplugs into ear canals,” *Annals of occupational hygiene*, vol. 46, no. 6, pp. 527–530, 2002.
- [105] W. J. Murphy, M. R. Stephenson, D. C. Byrne, B. Witt, and J. Duran, “Effects of training on hearing protector attenuation,” *Noise and Health*, vol. 13, no. 51, p. 132, 2011.
- [106] A. G. Samelli, C. H. Rocha, P. Theodósio, R. R. Moreira, and I. F. Neves-Lobo, “Training on hearing protector insertion improves noise attenuation,” in *CoDAS*, vol. 27, pp. 514–519, SciELO Brasil, 2015.
- [107] V. Zimpfer, G. Andeol, Y. Demumieux, A. Job, P. Hamery, G. Blanck, and S. De Mezzo, “Method to choose the ideal size of earplug,” *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3813–3813, 2017.

BIBLIOGRAPHY

- [108] T. C. Morata, C. L. Themann, R. F. Randolph, B. L. Verbsky, D. C. Byrne, and E. R. Reeves, “Working in noise with a hearing loss: perceptions from workers, supervisors, and hearing conservation program managers,” *Ear and hearing*, vol. 26, no. 6, pp. 529–545, 2005.
- [109] C. Giguère, C. Laroche, V. Vaillancourt, and S. D. Soli, “Modelling speech intelligibility in the noisy workplace for normal-hearing and hearing-impaired listeners using hearing protectors,” *International Journal of Acoustics and Vibration*, vol. 15, no. 4, p. 156, 2010.
- [110] P. N. Plyler and M. L. Klumpp, “Communication in noise with acoustic and electronic hearing protection devices,” *Journal of the American Academy of Audiology*, vol. 14, no. 5, pp. 260–268, 2003.
- [111] J. B. Tufts and T. Frank, “Speech production in noise with and without hearing protection,” *The Journal of the Acoustical Society of America*, vol. 114, no. 2, pp. 1069–1080, 2003.
- [112] C. J. Smalt, P. T. Calamia, A. P. Dumas, J. P. Perricone, T. Patel, J. Bobrow, P. P. Collins, M. L. Markey, and T. F. Quatieri, “The effect of hearing-protection devices on auditory situational awareness and listening effort,” *Ear and hearing*, vol. 41, no. 1, pp. 82–94, 2020.
- [113] J. Candido Fernandes, “Effects of hearing protector devices on speech intelligibility,” *Applied Acoustics*, vol. 64, no. 6, pp. 581–590, 2003.
- [114] S. M. Abel, A. Nakashima, and D. Saunders, “Speech understanding in noise with integrated in-ear and muff-style hearing protection systems,” *Noise and Health*, vol. 13, no. 55, p. 378, 2011.
- [115] A. Nakashima, S. M. Abel, and I. Smith, “Communication in military environments: influence of noise, hearing protection and language proficiency,” *Applied Acoustics*, vol. 131, pp. 38–44, 2018.
- [116] A. Bockstael, B. De Coensel, D. Botteldooren, W. D’Haenens, H. Keppler, L. Maes, B. Philips, F. Swinnen, and V. Bart, “Speech recognition in noise with active and passive hearing protectors: A comparative study,” *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3702–3715, 2011.
- [117] N. L. Vause and D. W. Grantham, “Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane,” *Human Factors*, vol. 41, pp. 282–294, June 1999.

BIBLIOGRAPHY

- [118] B. D. Simpson, R. S. Bolia, R. L. McKinley, and D. S. Brungart, "Sound localization with hearing protectors: Performance and head motion analysis in a visual search task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, pp. 1618–1622, SAGE Publications Sage CA: Los Angeles, CA, 2002.
- [119] W. G. Noble and G. Russell, "Theoretical and practical implications of the effects of hearing protection devices on localization ability," *Acta oto-laryngologica*, vol. 74, no. 1-6, pp. 29–36, 1972.
- [120] R. Mlynski and E. Kozłowski, "Localization of vehicle back-up alarms by users of level-dependent hearing protectors under industrial noise conditions generated at a forge," *International journal of environmental research and public health*, vol. 16, no. 3, p. 394, 2019.
- [121] B. D. Simpson, R. S. Bolia, R. L. McKinley, and D. S. Brungard, "The impact of hearing protection on sound localization and orienting behavior," tech. rep., Air Force research Laboratory, June 2005.
- [122] W. K. Vos, A. W. Bronkhorst, and J. A. Verhave, "Electronic pass-through hearing protection and directional hearing restoration integrated in a helmet," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3163–3163, 2008.
- [123] K. B. Lukas and W. A. Ahroon, "Free-field sound localization with nonlinear hearing protection devices," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 3080–3081, 2006.
- [124] C. Laroche, V. Vaillancourt, C. Giguère, M. Bibeau, V. Carroll, E. Gula, F. Nassrallah, H. Nélisse, and J. Boutin, "Effect of personal safety equipment (hearing protection and helmet) on the localization of reverse alarms," in *12th ICBEN Congress on Noise as a Public Health Problem*, pp. 1–8, 2017.
- [125] M. Takimoto, T. Nishino, K. Itou, and K. Takeda, "Sound localization under conditions of covered ears on the horizontal plane," *Acoustical Science and Technology*, vol. 28, no. 5, pp. 335–342, 2007.
- [126] V. Zimpfer and D. Sarafian, "Sound localization performance with the hearing protectors," in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, pp. 1–5, Acoustical Society of America, 2013.

BIBLIOGRAPHY

- [127] A. D. Brown, B. T. Beemer, N. T. Greene, T. Argo IV, G. D. Meegan, and D. J. Tollin, “Effects of active and passive hearing protection devices on sound source localization, speech recognition, and tone detection,” *PLoS One*, vol. 10, no. 8, pp. 1–21, 2015.
- [128] D. S. Brungart, A. J. Kordik, and B. D. Simpson, “The effects of single and double hearing protection on the localization and segregation of spatially-separated speech signals (L),” *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 1897–1900, 2004.
- [129] T. Joubaud, V. Zimpfer, A. Garcia, and C. Langrenne, “Degradation of front-back spectral cues induced by tactical communication and protective systems,” in *Proceedings Euronoise Conference, Maastricht, The Netherlands*, pp. 1943–1947, 2015.
- [130] R. R. Davis, W. J. Murphy, D. C. Byrne, and P. B. Shaw, “Acceptance of a semi-custom hearing protector by manufacturing workers,” *Journal of occupational and environmental hygiene*, vol. 8, no. 12, pp. 125–130, 2011.
- [131] B. Sheffield, D. Brungart, J. Tufts, and J. Ness, “The effects of elevated hearing thresholds on performance in a paintball simulation of individual dismounted combat,” *International journal of audiology*, vol. 56, no. sup1, pp. 34–40, 2017.
- [132] P. Rubak and L. G. Johansen, “Active hearing protector with improved localization performance,” in *Proceedings of InterNoise 1999, Lauderdale, FL, USA*, vol. 1, pp. 627–632, 1999.
- [133] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, “Measurement of a head-related transfer function database with high spatial resolution,” in *7th Forum Acusticum(EAA)*, pp. 1–6, Sept. 2014.
- [134] Austrian Academy of Sciences - Acoustics Research Institute, “Ari hrtf database.” https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=article&id=608&Itemid=857&lang=en. Accessed on May 12, 2020.
- [135] R. Bomhardt, M. de la Fuente Klein, and J. Fels, “A high-resolution head-related transfer function and three-dimensional ear model database,” in *Proceedings of Meetings on Acoustics 172ASA*, vol. 29, pp. 1–11, Acoustical Society of America, 2016.

BIBLIOGRAPHY

- [136] Austrian Academy of Sciences - Acoustics Research Institute, “sofaconventions.org.” [https://www.sofaconventions.org/mediawiki/index.php/SOFA_\(Spatially_Oriented_Format_for_Acoustics\)](https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)). Accessed on May 12, 2020.
- [137] P. Majdak and M. Noisternig, “AES69-2015: AES standard for file exchange - Spatial acoustic data file format,” *Audio Engineering Society*. (March 2015), 2015.
- [138] C. Guezenoc and R. Segquier, “HRTF individualization: A survey,” in *Audio Engineering Society Convention 145*, pp. 1–10, Oct. 2018.
- [139] International Organization for Standardization, “ISO 26101:2017 – Acoustics – Test methods for the qualification of free-field environments,” Apr. 2017.
- [140] B. Gardner and K. Martin, “HRTF measurements of a KEMAR dummy head microphone,” technical report, MIT Media Lab Perceptual Computing, May 1994.
- [141] JBL Professional, *Control1 Pro*, 2014.
- [142] P. Majdak, M. J. Goupell, and B. Laback, “3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training,” *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.
- [143] T. Hirahara, H. Sagara, I. Toshima, and M. Otani, “Head movement during head-related transfer function measurements,” *Acoustical science and technology*, vol. 31, no. 2, pp. 165–171, 2010.
- [144] M. Rothbucher, P. Paukner, M. Stimpfl, and K. Diepold, “The TUM-LDV HRTF database,” tech. rep., Technische Universität München, Lehrstuhl für Datenverarbeitung, 2014.
- [145] F. Denk, S. M. A. Ernst, S. D. Ewert, and B. Kollmeier, “Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles,” *Trends in hearing*, vol. 22, pp. 1–19, 2018.
- [146] G. Parmentier, A. Dancer, K. Buck, G. Kronenberger, and C. Beck, “Artificial head (ATF) for evaluation of hearing protectors,” *Acta Acustica United with Acustica*, vol. 86, no. 5, pp. 847–852, 2000.

- [147] International Telecommunication Union, “Series p: Terminals and subjective and objective assessment methods for objective measuring apparatus for artificial ears.” <https://www.itu.int/rec/T-REC-P.57-201112-I/en>, Dec. 2011. Accessed on May 19, 2020.
- [148] S. Müller and P. Massarani, “Transfer-function measurement with sweeps,” *Journal of the Audio Engineering Society*, vol. 49, pp. 443–471, June 2001.
- [149] F. Denk, S. D. Ewert, and B. Kollmeier, “Spectral directional cues captured by hearing device microphones in individual human ears,” *The Journal of the Acoustical Society of America*, vol. 144, no. 4, pp. 2072–2087, 2018.
- [150] C. I. Cheng and G. H. Wakefield, “Spatial frequency response surfaces: an alternative visualization tool for head-related transfer functions (hrtfs),” in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, vol. 2, pp. 961–964, IEEE, 1999.
- [151] M. Otani, Y. Iwaya, Y. Suzuki, and K. Itoh, “Numerical analysis of HRTF spectral characteristics based on sound pressures on a pinna surface,” in *Proceedings of 20th International Congress on Acoustics*, pp. 1–8, 2010.
- [152] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno, “Reconstructing head models from photographs for individualized 2D-audio processing,” in *Pacific Conference on Computer Graphics and Applications (Pacific Graphics)*, pp. 1719–1727, Oct. 2007.
- [153] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, “Mechanism for generating peaks and notches of head-related transfer functions in the median plane,” *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3832–3841, 2012.
- [154] L. Bonacina, A. Canciani, F. Antonacci, M. Marcon, A. Sarti, and S. Tubaro, “A low-cost solution to 3D pinna modeling for HRTF prediction,” in *2016 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2016. Proceedings. (ICASSP’16).*, pp. 301–305, IEEE, 2016.
- [155] A. Reichinger, P. Majdak, R. Sablatnig, and S. Maierhofer, “Evaluation of methods for optical 3-d scanning of human pinnas,” in *2013 International Conference on 3D Vision-3DV 2013*, pp. 390–397, IEEE, 2013.

BIBLIOGRAPHY

- [156] H. Braren and J. Fels, “A high-resolution individual 3D adult head and torso model for HRTF simulation and validation: 3D data,” tech. rep., Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Germany, 2020.
- [157] M. R. Bai and T.-C. Tsao, “Numerical modeling of head-related transfer functions using the boundary source representation,” *Journal of Vibration and Acoustics*, vol. 128, pp. 594–603, Oct. 2006.
- [158] 3D Systems, Inc., *Sense 3D scanner – Capture your world in 3D*, 2017.
- [159] H. Ziegelwanger, W. Kreuzer, and P. Majdak, “Mesh2HRTF: open-source software package for the numerical calculation of head-related transfer functions,” in *22st International Congress on Sound and Vibration*, pp. 1–8, 2015.
- [160] Y. Kahana, “Numerical modelling of the head-related transfer function.” University of Southampton, Nov. 2000. Lecture notes.
- [161] L. Greengard, *The rapid evaluation of potential fields in particle systems*. MIT press, 1988.
- [162] K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality*, vol. 1. Springer Nature Singapore, 2019.
- [163] A. Scharine, “The impact of helmet design on sound detection and localization,” *The Journal of the Acoustical Society of America*, vol. 117, p. 2561, 2005.
- [164] M. B. Gardner and R. S. Gardner, “Problem of localization in the median plane: Effect of pinnae cavity occlusion,” *The Journal of the Acoustical Society of America*, vol. 43, no. 2, pp. 400–408, 1973.
- [165] V. Best, S. Kalluri, S. McLachlan, S. Valentine, B. Edwards, and S. Carlile, “A comparison of CIC and BTE hearing aids for three-dimensional localization of speech,” *International Journal of Audiology*, vol. 49, pp. 723–732, 2010.
- [166] T. Joubaud, *Objective prediction of the effect of tactical communication and protective systems on sound localization performance*. PhD thesis, École Doctorale du Conservatoire National des Arts et Métiers, 2017.

BIBLIOGRAPHY

- [167] S. M. Abel, C. Burrell, and D. Saunders, “The effect of integrated hearing protection surround levels on sound localization,” tech. rep., DRDC Toronto Research Centre Toronto, Ontario Canada, 2015.
- [168] S. M. Abel, S. Boyne, and H. Roesler-Mulrone, “Sound localization with an army helmet worn in combination with an in-ear advanced communications system,” *Noise and Health*, vol. 11, no. 45, pp. 199–205, 2009.
- [169] T. van den Bogaert, E. Carette, and J. Wouters, “Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna,” *International Journal of Audiology*, vol. 50, no. 3, pp. 164–176, 2011.
- [170] C. T. Jin, A. Corderoy, S. Carlile, and A. van Schaik, “Spectral cues in human sound localization,” in *Advances in Neural Information Processing Systems*, pp. 768–774, 2000.
- [171] B. Seeber, *Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode*. PhD thesis, Lehrstuhl für Mensch-Maschine-Kommunikation der Technischen Universität München, 2003.
- [172] D. Byrne and W. Noble, “Optimizing sound localization with hearing aids,” *Trends in Amplification*, vol. 3, no. 2, pp. 51–73, 1998.
- [173] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *Journal of the audio engineering society*, vol. 45, no. 6, pp. 456–466, 1997.
- [174] 3M France, Département Solutions pour la Protection Individuelle, 3M France, Boulevard de l’Oise, 95006 Cergy-Pontoise Cedex, France, *Casques antibruit 3M™ Pelto™ Série X*, Nov. 2012. In French.
- [175] Honeywell Hearing Technologies AS, Sluppenvegen 12E, 7037 Trondheim, Norway, *QuietPro QP100*, Sept. 2016.
- [176] A. Nakashima, S. Sarray, and N. Fink, “Insertion loss of hearing protection devices for military impulse noise,” *Canadian Acoustics*, vol. 45, no. 3, pp. 148–149, 2017.
- [177] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster analysis*. John Wiley, 5 ed., 2011.
- [178] G. v. Békésy, “A new audiometer,” *Acta Oto-Laryngologica*, vol. 35, no. 5-6, pp. 411–422, 1947.

BIBLIOGRAPHY

- [179] Natus Medical Denmark ApS., *MADSEN Astera² Reference Manual*, Mar. 2019. (retrieved December, 29 2020).
- [180] B. Truax, ed., *Handbook for acoustic ecology*. World Soundscape Project, Simon Fraser University, and ARC Publications, second ed., 1999.
- [181] D. S. Brungart, B. W. Hobbs, and J. T. Hamil, “A comparison of acoustic and psychoacoustic measurements of pass-through hearing protection devices,” in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 70–73, IEEE, 2007.
- [182] R. S. Bolia, W. R. D’Angelo, P. J. Mishler, and L. J. Morris, “Effects of hearing protectors on auditory localization in azimuth and elevation,” *Human Factors*, vol. 43, no. 1, pp. 122–128, 2001.
- [183] V. Zimpfer, D. Sarafian, K. Buck, and P. Hamery, “Spatial localization of sounds with hearing protection devices allowing speech communication,” in *Proceedings of the Acoustics 2012 Nantes Conference*, pp. 527–532, 2012.
- [184] R. Pelzer, M. Dinakaran, F. Brinkmann, S. Lepa, P. Grosche, and S. Weinzierl, “Head-related transfer function recommendation based on perceptual similarities and anthropometric features,” *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3809–3817, 2020.
- [185] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, “Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data,” in *Audio Engineering Society Convention 123*, Oct. 2007.
- [186] H. S. Braren and J. Fels, “A High-Resolution Head-Related Transfer Function Data Set and 3D-Scan of KEMAR,” tech. rep., Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Germany, 2020.
- [187] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, “Structural composition and decomposition of HRTFs,” in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 103–106, IEEE, 2001.

BIBLIOGRAPHY

- [188] E. A. G. Shaw, *Binaural and Spatial Hearing in Real and Virtual Environments*, vol. 25, ch. Acoustical features of the human external ear, pp. 25–47. Mahwah, NJ: Lawrence Erlbaum, 1997.
- [189] J.-P. Barral and A. Croibier, “Chapter 25 - Ear,” in *Manual Therapy for the Cranial Nerves* (J.-P. Barral and A. Croibier, eds.), ch. 25, pp. 227–238, Edinburgh: Churchill Livingstone, 2009.
- [190] U.S. Food and Drug Administration, “Types of hearing aids.” <https://www.fda.gov/medical-devices/hearing-aids/types-hearing-aids>, Jan. 2018. (retrieved May 17, 2019).
- [191] National Health Service, “Hearing aids.” <https://www.nhs.uk/live-well/healthy-body/hearing-aids/>, Sept. 2020. (retrieved January 28, 2021).
- [192] F. Denk, S. D. Ewert, and B. Kollmeier, “On the limitations of sound localization with hearing devices,” *The Journal of the Acoustical Society of America*, vol. 146, no. 3, pp. 1732–1744, 2019.
- [193] V. Durin, S. Carlile, P. Guillon, V. Best, and S. Kalluri, “Acoustic analysis of the directional information captured by five different hearing aid styles,” *The Journal of the Acoustical Society of America*, vol. 136, pp. 818–828, Aug. 2014.
- [194] U. Zölzer, *Digital audio signal processing*, vol. 9. Wiley Online Library, 2008.
- [195] U. Zölzer, *DAFX: digital audio effects*. John Wiley & Sons, second ed., 2011.
- [196] J. S. Abel and D. P. Berners, “Filter design using second-order peaking and shelving sections,” in *ICMC*, 2004.
- [197] K. I. McAnally and R. L. Martin, “Sound localization with head movement: implications for 3-d audio displays,” *Frontiers in neuroscience*, vol. 8, p. 210, 2014.
- [198] J. C. Makous and J. C. Middlebrooks, “Two-dimensional sound localization by human listeners,” *The journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990.
- [199] V. Pulkki and T. Hirvonen, “Localization of virtual sources in multichannel audio reproduction,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 105–119, Jan. 2005.
- [200] A. Kerstenetzky, “PolyJet™ Materials Reference Guide,” tech. rep., Stratasys, Aug. 2020.

- [201] F. Zotter and M. Frank, “All-round ambisonic panning and decoding,” *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.
- [202] P. W. Gillett, *Head mounted microphone arrays*. PhD thesis, Virginia Tech, 2009.
- [203] S. Hengy, “Sniper localization using a helmet array,” tech. rep., Institut Franco-Allemand de recherches de Saint-Louis, 2006.
- [204] W. L. Chapin, A. R. Jost, B. A. Cook, F. Surucu, S. Foster, M. Bolas, I. McDowall, E. R. Lorimer, P. M. Zurek, J. G. Desloge, R. E. Beaudoin, B. Shinn-Cunningham, and N. Durlach, “Concept and technology exploration for transparent hearing,” tech. rep., U.S. Government Department of Defense Agencies, 2003.
- [205] B. U. Seeber, “Weighting of binaural cues in the absence of a reflection,” in *Fortschritte der Akustik – DAGA 2010*, pp. 641–642, 2010.
- [206] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, “Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint,” *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.
- [207] E. M. Benjamin, R. Lee, and A. J. Heller, “Localization in horizontal-only ambisonic systems,” in *Audio Engineering Society Convention 121*, pp. 1–13, Audio Engineering Society, Oct. 2006.
- [208] H. Pomberger, A. Sontacchi, M. Frank, T. Gmeiner, and M. Lucchi, “Improved localisation in the median plane with cue-preserving headphones,” in *DAGA 2018 - 44. Jahrestagung für Akustik, Munich, Germany*, pp. 948–951, 2018.
- [209] F. Zotter and M. Frank, *Ambisonics*, vol. 19 of *Springer Topics in Signal Processing*. Springer International Publishing, first ed., 2019.
- [210] VVAudio, “VVMic™ for TetraMic™ Ver. 3.5.1.” <http://www.vvaudio.com/downloads>, June 2007. (retrieved February 1, 2021).

Appendix A

Résumé étendu des travaux de thèse

A.1 Chapitre 1

L'humain perçoit des ondes acoustiques à partir d'une amplitude de $20 \mu\text{Pa}$ ($\cong 0 \text{ dB SPL}$) entre 20 Hz et 20 kHz [1, 3]. La sensibilité du système auditif dépend de la fréquence [17]. Le seuil de douleur apparaît à des niveaux sonores compris entre 20 Pa ($\cong 120 \text{ dB SPL}$) et 200 Pa ($\cong 140 \text{ dB SPL}$) [4]. Des courbes de pondération, basées sur l'inverse du seuil de perception, permettent d'estimer le niveau sonore perçu d'un son large bande [25]. Pour évaluer l'exposition à des bruits de durées limitées, un niveau de bruit équivalent à un bruit constant est déterminé sur l'équivalent journée de travail ($\cong 8 \text{ h}$) [26].

Les sons environnementaux sont localisés grâce à la différence de temps interaurale (Interaural Time Difference (ITD)), la différence de niveau interaurale (Interaural Level Difference (ILD)) et les indices spectraux. La considération de l'ITD et ILD permet de décrire un cône de confusion dans lequel la source sonore peut être placée. À l'aide des indices spectraux, la position exacte sur le cône de confusion est déterminée. Les indices spectraux sont uniques pour les différents angles d'azimut et d'élévation, mais aussi individuels pour chaque humain. Ces indices sont introduits par l'oreille externe et définis par son anatomie. Le cerveau est bien adapté aux indices spectraux individuels. L'ITD, l'ILD et les indices spectraux décrivent la fonction de transfert relative à la tête (Head-Related Transfer Function (HRTF)). Les HRTFs fournissent des informations spatiales sur les différentes sources acoustiques, permettant de séparer les sources individuelles (c.f. "cocktail party effect" [27]) et de les localiser. Dans des environnements virtuels, on se sert des HRTFs pour créer des scènes acoustiques avec des sources placées virtuellement dans l'espace. Les changements de l'HRTF individuelle, provoqués en

insérant des bouchons ou en utilisant des casques de protection réduisent fortement la localisation sonore [51, 50].

Dans des situations avec un niveau de bruit très fort, il faut utiliser les protecteurs auditifs (Hearing Protection Devices (HPDs)) pour se protéger des risques possibles, à la fois des maladies indirectement liées à l'oreille (maladies cardio-vasculaires, stress, insomnies) et des maladies de l'oreille comme la surdit  [88, 87, 89]. Par contre, porter des HPDs emp che la conversation face-  face [108, 109], mais aussi la localisation fiable des sons environnementaux [50, 117, 118, 119, 120, 121]. Souvent, les utilisateurs d cident de pouvoir communiquer et d'analyser leur environnement correctement et donc, enl vent la protection acoustique en prenant des risques sanitaires. Les  tudes pr sent es dans la suite ont pour objectif d'orienter la conception d'HPDs qui permettraient une localisation des sons environnementaux et donc augmenteraient l'acceptation des HPD en r duisant de ce fait les cas de perte de l'audition.

Pour tester les solutions propos es un **environnement acoustique virtuel** (Virtual Acoustic Environment (VAE)) a  t  pens  planifi  et mis en place en deux phases. Ce VAE consiste en 16 haut-parleurs professionnels, espac s de 22.5° sur un cercle de 2.20 m de diam tre, c.f. Figure A.1. Un logiciel de guidage est d velopp  et g re le positionnement des sources sonores. Gr ce   la m thode de "vector base amplitude panning", le son peut  tre positionn  en continu sur tout l'ensemble du cercle. Le logiciel est  galement muni d'une interface utilisateur qui peut  tre utilis  pour des tests de localisation de sources sonores. Dans sa premi re phase, la sortie audio st r o de l'ordinateur est utilis e et la VAE ne permet pas de pr senter des sons dans plusieurs directions simultan es. Dans la deuxi me phase de d veloppement, une interface audio professionnelle est int gr e permettant le contr le direct sur le signal de chaque haut-parleur et donc rajoute la possibilit  de placer plusieurs sources sonores simultan es avec des positions ind pendantes.

A.2 Chapitre 2

Ce chapitre traite de l'interaction entre les HPDs et la localisation des sources sonores sous 5 **conditions d' coute**. Ces conditions d' coute sont d finies comme suit, o  HPD P0 est une condition d' coute naturelle et HPD P1   P4 sont des conditions d' coute non-naturelles.

- HPD P0 : l' coute naturel sans HPD.

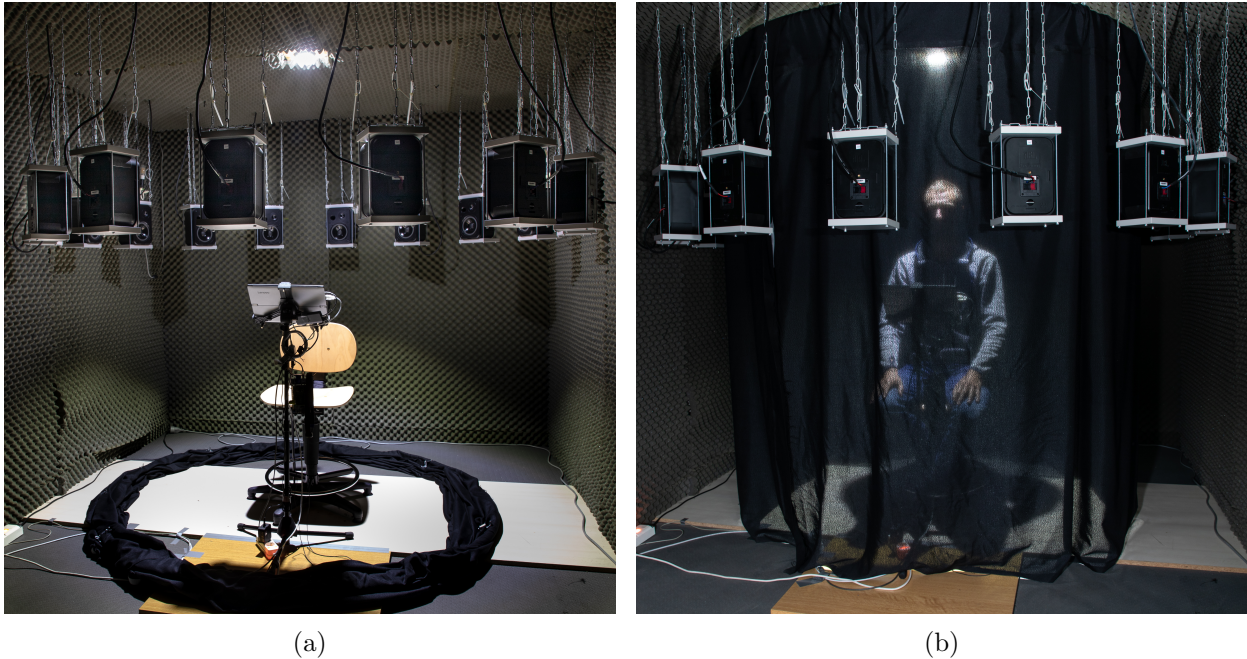


FIGURE A.1 : VAE à 16 haut-parleurs positionnés dans un cercle de 2.20 m de diamètre. La position d’écoute est au milieu du cercle, les haut-parleurs sont cachés derrière un rideau acoustique. Gauche : le rideau est posé par terre. Droite : le rideau est mis en place.

- HPD P1 : l’écoute avec l’HPD *ISL* “Bang” qui est un bouchon actif avec un embout d’oreille à ailette universel.

HPD P1C : équivalent à l’HPD P1 mais avec un bouchon d’oreille personnalisé et moulé, disponible pour les têtes artificielles Harry34 et Harry33.

- HPD P2 : l’écoute avec l’HPD *ZTac* “Z111” qui est un casques auditif actif.
- HPD P3 : l’écoute avec l’HPD *Nacre* “QuietPro” qui est un bouchon actif avec un embout d’oreille en mousse.
- HPD P4 : l’écoute avec l’HPD *3M* “X5A” qui est un casque auditif passif.

Dans la chambre anechoïque, on mesure les HRTFs de trois têtes artificielles avec chaque condition d’écoute (HPD P0 à P4). Les têtes artificielles sont :

- Harry34 : tête artificielle du French-German Research Institute of Saint-Louis (ISL), équipée avec des simulateurs d’oreille simplifiée de type 3.4 conforme à l’ITU-T P.57 [147].

- Harry33 : tête artificielle de l'ISL, équipée avec des simulateurs d'oreille avancée de type 3.3 conforme à l'ITU-T P.57 [147].
- HATS33 : tête artificielle de *B&K* Type 5128 "Head and Torso Simulator".

Le signal de référence est un sinus glissant et la distance entre le haut-parleur et la position de la tête est de 3 m. Cette distance est limitée par la longueur de la chambre anechoïque de 5.60 m. La condition d'écoute HPD P0 est mesurée deux fois pour chaque tête. On obtient 20 HRTFs, dont la résolution angulaire est de 22.5° et la plage de fréquences est de 16 Hz à 25.6 kHz avec 1600 points espacés logarithmiquement. Les signaux mesurés sont convertis pour obtenir la Directional Transfer Function (DTF) qui est, dans la suite, aussi appelée HRTF.

L'analyse des **modifications d'HRTF** introduites par les protections est basée sur la racine-carrée de l'erreur quadratique moyenne (Root Mean Square Error (RMSE)) entre les conditions d'écoute non-naturelles (HPD P1 à P4) et la condition d'écoute naturelle (HPD P0), c.f. Équations (A.1) et (A.2). L'HRTF $_{P_i}$ correspond à l'HRTF avec une écoute non-naturelle avec $i = 1, \dots, 4$, et HRTF $_0$ correspond à l'HRTF avec l'écoute naturelle.

$$E(f, \phi) = 20 \cdot \log_{10} \left(\left| \frac{\text{HRTF}_{P_i}(f, \phi)}{\text{HRTF}_0(f, \phi)} \right| \right) \quad (\text{A.1})$$

$$D(f) = \sqrt{\frac{1}{16} \cdot \sum_{\phi=0^\circ}^{337.5^\circ} E(f, \phi)^2} \quad (\text{A.2})$$

La Figure A.2, présente les valeurs RMSE par bandes de fréquences établies par Zwicker [12]. Tout d'abord, on s'aperçoit que les courbes diminuent très rapidement dans les trois premières bandes, puis qu'elles augmentent avec des pentes différentes et arrivent sur des maximums vers la bande N° 24. La pente avec laquelle les courbes remontent dépend fortement de l'HPD et non de la tête artificielle. Pour une comparaison entre les bandes N° 7 et N° 20, P1 et P3 (type de bouchon) possèdent une RMSE beaucoup moins grande qu'entre P2 et P4 (type de casque). Sur une bande fréquentielle très large (jusqu'à la bande N° 15), les bouchons introduisent un changement inférieur à 2 dB dans les HRTFs, alors que le changement dépasse 2 dB avec des casques à partir de la bande N° 5. Dans les hautes fréquences comme dans les basses fréquences, la bande passante du haut-parleur (d'après le

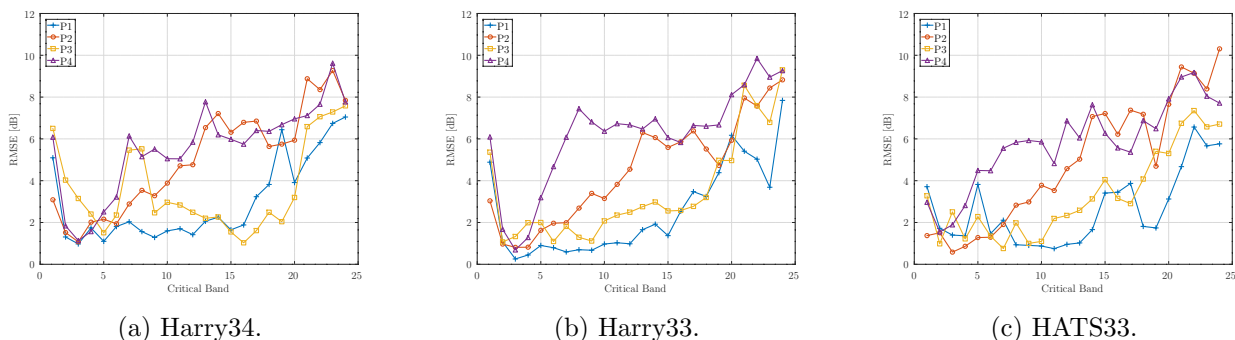


FIGURE A.2 : RMSE entre des paires d’écoutes naturelles (P0) et non-naturelles (P1, P2, P3, et P4) par bandes de Zwicker (“critical bands”) [12] pour les trois têtes artificielles.

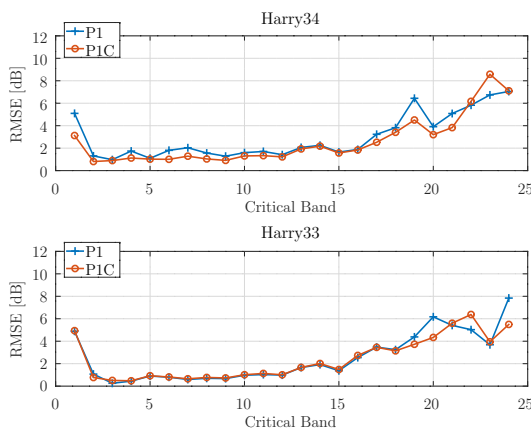


FIGURE A.3 : RMSE entre des paires d’écoutes naturelles (P0) et non-naturelles (P1 et P1C) par bandes de Zwicker (“critical bands”) [12] pour les deux têtes artificielles.

fournisseur de 100 Hz à 18 kHz) limite l’énergie du signal de référence, réduisant la pertinence des résultats.

La Figure A.3 compare les RMSE dûs aux différents types de bouchons d’oreilles (universels ou personnalisés). Pour les deux têtes artificielles, on obtient des courbes presque identiques. Sauf aux hautes fréquences (supérieures à la bande N° 17), les bouchons d’oreilles montrent des différences significatives, liées au positionnement du bouchon dans l’oreille et à la bande passante du haut-parleur.

Avec un **test de localisation**, on évalue la performance de localisation des sources sonores sous les conditions d’écoute HPD P0 à P4. Le test est effectué dans la VAE à 16 haut-parleurs. Pendant ce test à choix non-forcé, on présente un bruit blanc de 200 ms et on demande aux sujets d’indiquer sur une tablette tactile la direction dans laquelle ils ont perçu le son. On utilise les directions $\phi_t \in$

$\{0^\circ, 22.5^\circ, \dots, 337.5^\circ\}$ et chaque direction est testée 5 fois, soit 80 sons à écouter par condition d'écoute. La moitié des sujets (groupe de test) passe une phase d'entraînement à 32 sons avec chaque condition d'écoute. On commence toujours avec l'écoute naturelle, suivie, dans un ordre aléatoire, des conditions d'écoute non-naturelles. Avec un test de Békésy on vérifie le seuil d'audition des sujets aux fréquences d'octave entre 125 Hz et 8 kHz.

40 sujets avec une audition normale et un âge moyen de $\mu = 31.45$ ans ($\sigma = 10.47$) ont participé à ce test de localisation. Le temps de réponse est en moyenne de 2.42 s et diminue logarithmiquement au cours du test de 3.71 s au début vers 2 s à la fin. Ceci implique que le temps de réponse avec HPD P0 (écoute naturelle) est toujours plus élevé qu'avec les conditions non-naturelles. La perte de motivation et l'adaptation au test sont des raisons majeures pour la décroissance du temps de réponse.

La matrice de confusion de la Figure A.4a montre avec une résolution de $\mp 5.625^\circ$ qu'il y a peu de confusions gauche-droite et mixtes (peu de points dans le sous-quadrant haut-gauche et bas-droite). Par contre, il y a un certain nombre de confusions avant-arrière (points sur les anti-diagonales). Une grosse partie des points se répartissent autour de la diagonale principale, ce que l'erreur angulaire non-signifiée de $|\epsilon| = 13.18^\circ$ confirme. L'évolution de $|\epsilon|$, visualisée dans la Figure A.4b, montre clairement l'influence d'un HPD sur la performance de localisation. En dépassant 20 % du test, c'est-à-dire en passant de HPD P0 aux HPDs non-naturels, $|\epsilon|$ augmente en moyenne instantanément de 2.53° soit de 22.67 %. Les confusions avant-arrière, c.f. Figure A.4c, dépendent significativement de la condition d'écoute (p-value < 0.001). Avec HPD P0, le taux de confusions est 0.052 (groupe de test) et 0.084 (groupe de contrôle). Cette valeur augmente à 0.28 pour les HPDs de type bouchon et à 0.39 pour les HPDs de type casque. Avec une précision d'environ 0.6, la plupart des réponses se trouvent dans un intervalle de $\mp 4.5^\circ$ autour de la position d'un haut-parleur, peu importe si c'est le bon haut-parleur ou non.

On a pu montrer que l'effet des HPDs sur les HRTFs est inversement proportionnel à la performance de localisation des sources sonores. Plus l'HRTF est détériorée, plus il y a de confusions avant-arrière. Concernant cette relation, les HPDs de type bouchon sont à préférer aux HPDs de type casque.

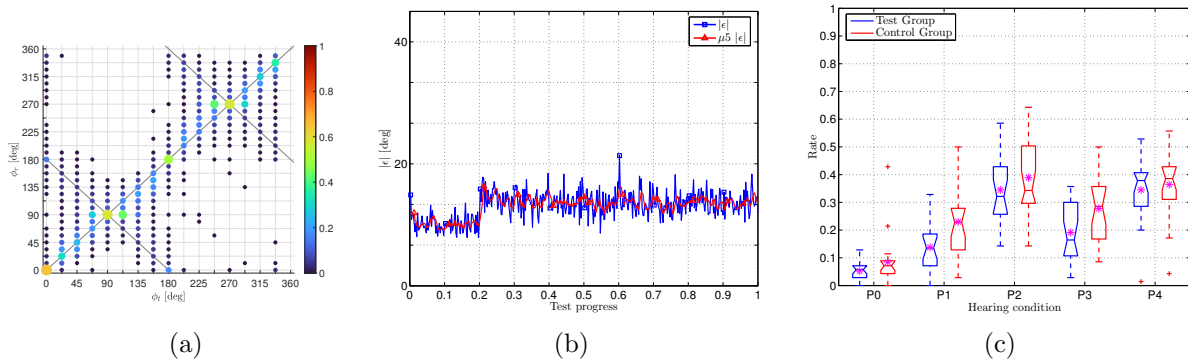


FIGURE A.4 : Résultats du test de localisation. La matrice de confusion dans la Figure A.4a donne une impression globale sur la performance de localisation. L'évolution de l'erreur angulaire non-signée $|\epsilon|$, moyennée sur toutes les conditions d'écoute au cours du test est montrée dans la Figure A.4b. Les confusions avant-arrière par groupe et par conditions d'écoute sont visualisées dans la Figure A.4c.

A.3 Chapitre 3

Dans ce chapitre on applique les **simulations d'HRTF**. Les têtes artificielles (Harry34, Harry33, et HATS33) sont numérisées avec un scanner 3D et des maillages 3D sont utilisés pour faire des simulations numériques, basées sur la méthode des éléments finis de frontière (Boundary Element Method (BEM)). Les simulations sont faites pour la même échelle de fréquences (1600 pas logarithmiques dans l'intervalle [16 Hz; 25.6 kHz]) et d'angles d'incidence ($\phi = \{0^\circ, 22.5^\circ, \dots, 337.5^\circ\}$) que les mesures du chapitre précédent.

La RMSE entre l'HRTF mesurée et simulée est de 2.54 dB pour Harry34, 2.10 dB pour Harry33 et 1.91 dB pour HATS33. Comparativement aux RMSEs entre deux têtes différentes qui varient entre 2.22 dB et 3.07 dB, on observe en particulier pour Harry34, une erreur entre la simulation et la mesure plus grande que entre deux têtes différentes. Cette observation est liée à la mesure d'HRTF de Harry34 qui a une répétabilité réduite. Le montage expérimental pour Harry34 sur la table tournante inclut un réhausseur qui alourdit et déséquilibre le dispositif, conduisant facilement à des erreurs de positions. Cet effet est également observé, mais de manière moins prononcée pour Harry33. Les RMSEs par bandes de Zwicker entre les HRTFs mesurées et simulées des trois têtes artificielles sont données dans la Figure A.5. À part pour la bande N° 1, l'erreur dépasse sérieusement la limite de 2 dB à partir de la bande N° 10 (Harry34 et Harry33) et N° 18 (HATS33). Pour Harry34 et Harry33, le problème de stabilité suite à la mise en place réduit la concordance par rapport à HATS33. Des écarts très fins dans la partie de l'oreille externe entre la tête artificielle réelle et son maillage 3D cause des résultats

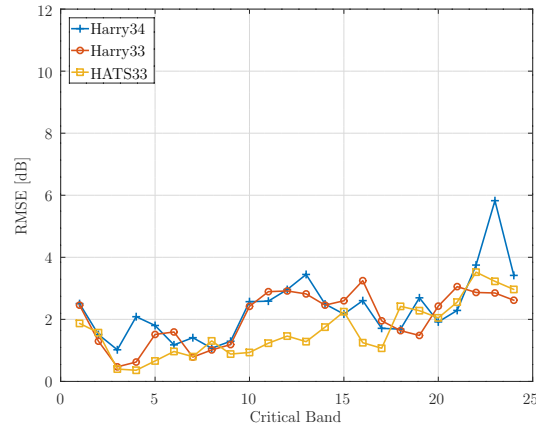
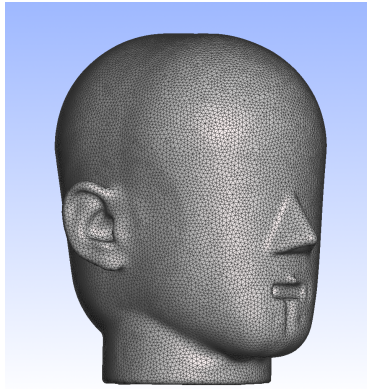


FIGURE A.5 : Les RMSEs par bandes de Zwicker entre les HRTFs mesurées et simulée des trois têtes artificielles.

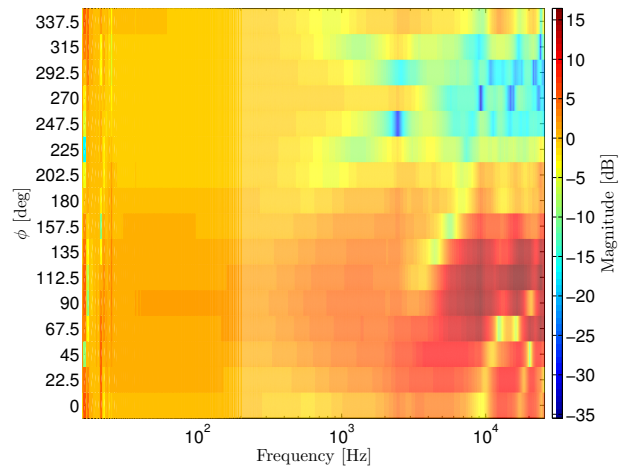
légèrement différents dans les hautes fréquences.

La **contribution de l'oreille externe** est, dans un premier temps, évaluée avec le maillage de la tête artificielle HATS33, c.f. Figure A.6a. À partir de ce dernier, un second modèle est créé en enlevant les oreilles externes et en ajoutant un conduit de forme cylindrique, c.f. Figure A.6c. Un dernier modèle consiste à extraire l'oreille externe du côté droit, c.f. Figure A.6e. Les HRTFs pour ces trois modèles sont montrés dans les Figures A.6b, A.6d et A.6f. La méthode du gradient conjugué (Conjugate-Gradient Solver (CGS)) ne converge pas dans les basses fréquences (< 30 Hz) pour les maillages HATS33 et HATS00, produisant des artefacts dans les HRTFs. En comparant les Figures A.6d et A.6f avec la Figure A.6b, on obtient que la tête seule (HATS00) introduit l'ILD et l'ITD, mais aussi des indices spectraux symétriques à l'axe interaural et donc non utilisables pour distinguer l'avant et l'arrière. Par contre, l'oreille externe introduit des indices spectraux supérieurs à 1 kHz asymétrique par rapport à l'axe interaural. Avec les indices spectraux de l'oreille externe la localisation des sons donne des résultats fiables. Pour les HPDs de type bouchon, cela signifie qu'ils doivent couvrir l'oreille externe le moins possible.

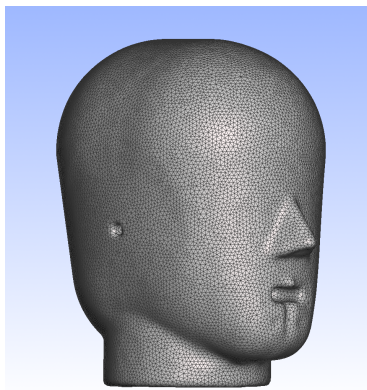
Pour réduire la durée d'acquisition d'une HRTF, on évalue l'idée de combiner une HRTF (générique) d'une tête sans oreille, avec l'HRTF des oreilles externes (individuelles). Deux méthodes pour la **recomposition de l'HRTF** à partir de l'HRTF de la tête sans oreille (HATS00) et l'HRTF de l'oreille externe (oreille externe de HATS33) sont présentées. La méthode 1 additionne les deux HRTFs, la méthode 2 moyenne les deux HRTFs. Les RMSEs entre les résultats de ces deux méthodes et l'HRTF



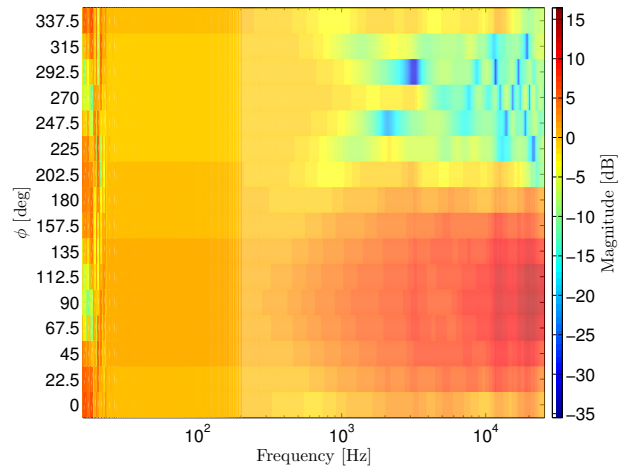
(a) HATS33.



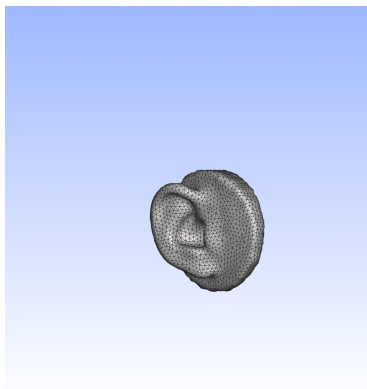
(b) HRTF de HATS33



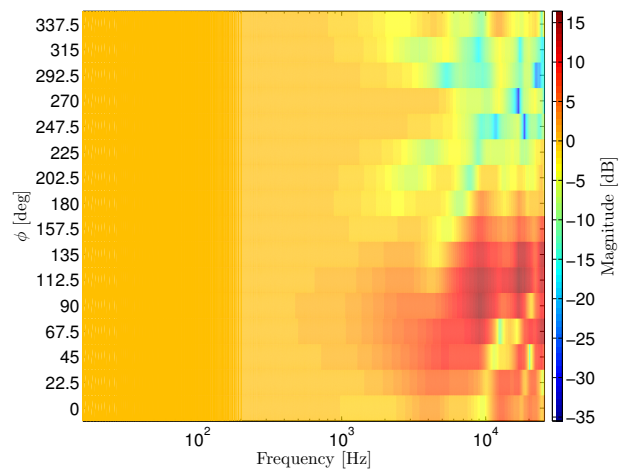
(c) HATS00.



(d) HRTF de HATS00.



(e) Oreille externe de HATS33.



(f) Pinna-Related Transfer Function (PRTF) de l'oreille externe de HATS33.

FIGURE A.6 : Maillages 3D (gauche) et résultats des simulations numériques d'HRTF (droite).

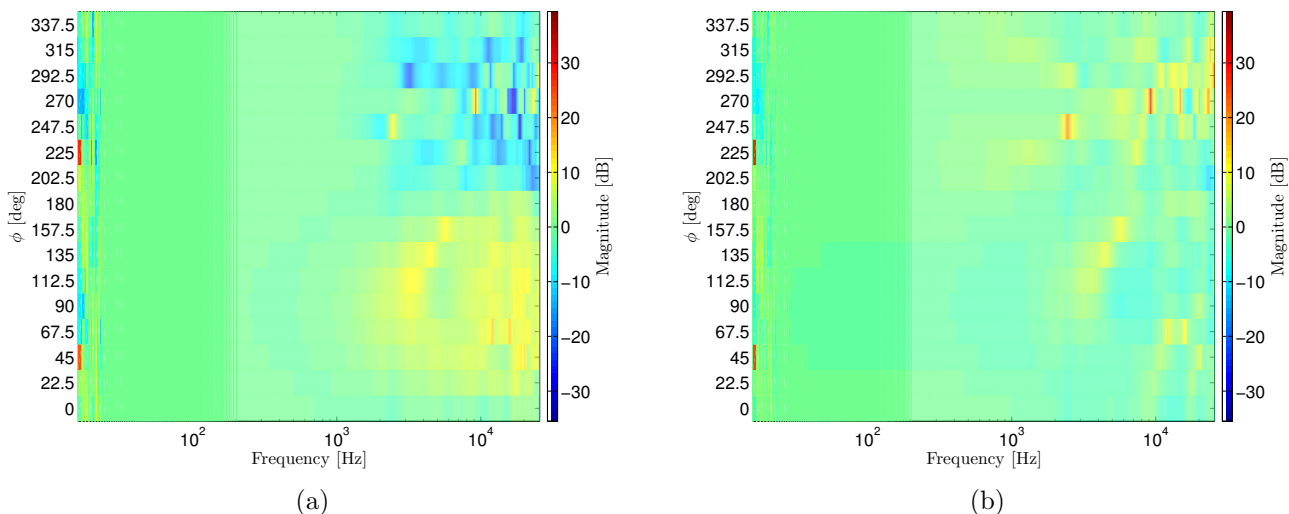


FIGURE A.7 : erreurs entre l’HRTF d’origine (HATS33) et les HRTFs composées, en utilisant méthode de sommation (Figure A.7a) et de moyenne (Figure A.7b). Les valeurs positives correspondent à une surestimation de l’HRTF d’origine.

d’origine (HATS33) sont montrées sur la Figure A.7. En dessous de 1 kHz, les deux méthodes donnent les mêmes erreurs. Par contre au dessus de 1 kHz, la méthode des sommes résulte dans une RMSE de 4.34 dB et la méthode des moyennes donne une RMSE de 2.68 dB. Néanmoins, prendre la moyenne réduit les distances entre les maxima et minima locaux qui définissent les indices spectraux individuels qui sont importants pour la localisation sonore. Il serait nécessaire de réaliser des études supplémentaires car ces deux méthodes sont identique à un gain près. Notamment, une évaluation subjective permettrait de mieux comprendre cet effet.

A.4 Chapitre 4

Dans une première partie un **modèle analytique d’HRTF**, contrôlé par l’angle horizontal ϕ , est développé. Ce modèle est basé sur l’analyse des différences spectrales entre l’avant et l’arrière de l’HRTF simulée et mesurée de Harry33. Avec cette analyse on voit que la plupart des différences spectrales entre l’avant et l’arrière sont localisées entre 2 kHz et 8 kHz. Cette plage de fréquences est très proche de la bande de Blauert entre 1.86 kHz et 7.03 kHz [40]. Le modèle se focalise donc sur la décade fréquentielle allant de 1 kHz à 10 kHz. Il est réalisé avec une superposition de 5 filtres peak d’ordre deux à réponse impulsionnelle infinie (Infinite Impulse Response (IIR)) [194]. Les fréquences centrales, les bandes passantes et les gains de ces filtres sont définis à partir d’une analyse des différences

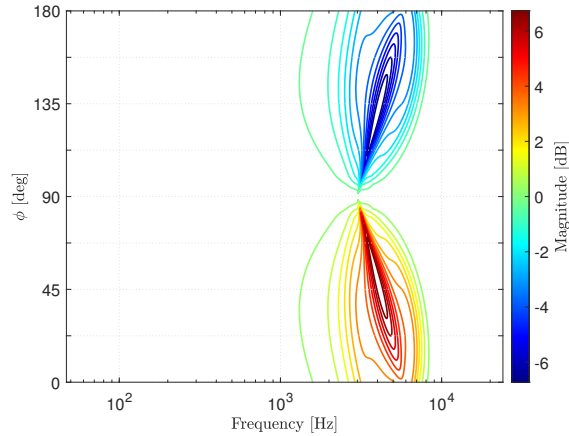


FIGURE A.8 : Les courbes de niveaux de l'amplitude de l'HRTF paramétrée.

spectrales avant-arrière des HRTFs naturelles des têtes artificielles. La fonction obtenue est montrée sur la Figure A.8. L'erreur entre le modèle développé et les HRTFs d'origine est inférieure ou égale à 1.80 dB (pour l'HRTF d'origine mesurée) et 1.31 dB (pour l'HRTF d'origine simulée).

Dans une deuxième partie on évalue avec un **test de localisation** un grand nombre de filtres numériques concernant la performance subjective de discrimination entre des sons placés virtuellement à l'avant et à l'arrière. L'ensemble des filtres inclut des filtres de shelving (3 kHz High-Shelf Filter (HSF) [76], le modèle analytique d'HRTF, les bandes de Blauert [40]) et des HRTFs. Deux méthodes de génération de signaux sont utilisées : par traitement ambisonique et par traitement discret. Dans la méthode ambisonique les signaux sont enregistrés dans le VAE avec un microphone ambisonique de premier ordre, et les filtres sont appliqués sur les signaux des haut-parleurs virtuels, avant obtenir le signal stéréo. Dans la méthode discrète, les signaux sont générés par l'application du filtre en fonction de la direction et un décalage entre le signal de gauche et de droite, qui correspond à l'ITD. Ce test de localisation est en deux parties. La première partie ne contient que des stimuli ambisoniques, 13 filtres distincts, 4 directions différentes ($\phi = \{0^\circ, 67.5^\circ, 180^\circ, 247.5^\circ\}$), et deux sons différents (voix masculine et bruit d'une cigale). La deuxième partie contient des stimuli distincts et ambisoniques, 4 filtres, 12 directions différentes ($\phi = \{22.5^\circ, 45^\circ, \dots, 337.5^\circ\}$ sans $\phi = \{90^\circ, 180^\circ, 270^\circ\}$), et qu'un seul son de base (voix masculine). Les participants écoutent les stimuli dans un espace acoustique à bruit réduit par un casque audio stéréo et indiquent leur réponse sur une tablette à écran tactile en dirigeant une flèche dans la direction perçue.

Le temps de réponse est en moyenne de 2.35 s (partie 1) et 2.40 s (partie 2) et diminue logarithmiquement avec l'expérience. Les données sur les directions perçues montrent qu'avec une résolution de 5.625° , les sujets répondent le plus souvent avec un erreur angulaire signée $\epsilon = \{0^\circ, \pm 22.5^\circ\}$ (partie 1) et $\epsilon = \{\pm 22.5^\circ, \pm 45^\circ, \pm 67.5^\circ\}$ (partie 2). Les sujets préfèrent répondre aux angles $\phi_r = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ (partie 1) et aux angles $\phi_r = \{90^\circ, 270^\circ\}$ (partie 2), ce qui signifie que la discrimination gauche-droite est maintenue, par contre la position exacte est perdue. La performance de distinction entre l'avant et l'arrière ne varie pas entre les filtres de la partie 1. On obtient un taux de confusion avant-arrière en moyenne de 31.43 %. Un test de Kruskal-Wallis montre qu'il n'y a pas de différences significatives entre ces filtres (p-value : 0.72). Le grand nombre de filtres avec des indices trop similaires ne permet pas aux participants de remarquer les fines nuances spectrales entre les sons. Les calculs pour déterminer le taux de confusion avant-arrière ne prennent pas en compte les sons pour $\phi_t = \{0^\circ, 180^\circ\}$, c'est-à-dire que pour cette première partie, seulement 50 % des données sont représentées par la confusion avant-arrière. Pour les filtres de la partie 2, le taux de confusion avant-arrière est de 43.78 % en moyenne et varie dans l'intervalle [36.25 %; 52.22 %]. Un test de Kruskal-Wallis montre une différence significative entre les filtres (p-value : $\ll 0.001$). Concernant les filtres en combinaison avec la méthode discrète, l'HRTEF non-individuelle donne les indices spectraux les moins fiables. Le HSF de 3 kHz, la bande de Blauert et le modèle analytique d'HRTEF proposent des indices plus compréhensibles aux participants. Par conséquent, en travaillant avec des filtres non-individuels il est beaucoup plus prometteur d'utiliser des filtres avec une réponse fréquentielle lisse, pas trop détaillée, plutôt que des filtres avec une réponse fréquentielle avec plein de détails. Les confusions gauche-droite et mixtes montrent des résultats similaires : pour les filtres en combinaison avec la méthode discrète, le taux ne dépasse pas 2.27 %, par contre en combinaison avec la méthode ambisonique le taux augmente jusqu'à 12.37 %. En utilisant le filtre d'HRTEF pour la méthode discrète et la méthode ambisonique, on observe avec cette deuxième partie du test d'écoute que la méthode ambisonique est moins avantageuse que la méthode discrète. La méthode ambisonique introduit une ILD très faible qui n'est plus correctement identifiée par les participants. Dans la partie 1 qui ne dispose que d'une ILD très faible, les sujets s'y sensibilisent, par contre dans la partie 2 qui dispose d'une ILD très faible (de la méthode ambisonique) mais aussi d'une ITD (de la méthode discrète), les sujets ne montrent pas cette sensibilisation. En plus, du fait que le test est à choix non-forcé, les sujets répondent avec une précision limitée ce qui augmente le taux de confusions. Les angles $\phi = \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ correspondent aux frontières des quadrants. Quand les

sujets veulent répondre vers une de ces directions, ils doivent positionner la flèche sur la tablette avec une bonne précision car une légère faute dirigerait le curseur dans le quadrant voisin, induisant une erreur avant-arrière, gauche-droite ou mixte.

A.5 Chapitre 5

On développe 4 prototypes pour des protections acoustiques avancées de type casque. Le but de leur design est de préserver des indices spectraux et donc la performance de localisation des sources sonores. Les coques des prototypes sont basées sur la coque de l'HPD P2 du Chapitre 2, c.f. Annexe A.2, en enlevant les ouvertures pour des boutons, donnant une géométrie très simple et symétrique par rapport au plan frontal. Les prototypes sont utilisés en tant que capteurs et ne disposent que de microphones. Le traitement du signal inclut pour chaque prototype l'égalisation des microphones, celui-ci est réalisé sur un ordinateur après avoir enregistré les signaux. Une fois traités, ils peuvent être écoutés avec un casque audio standard.

Le **Prototype A** est équipé de 6 microphones unidirectionnels (*Kingstate* KEIG4537TFL-N), dont 3 sont montés sur l'extérieur de chaque coque. Les microphones sont orientés vers l'avant, les côtés et l'arrière, numérotés de $M1$ à $M6$ dans le sens inverse des aiguilles d'une montre en commençant par le microphone avant-gauche, c.f. Figure A.9a. Des mesures de directivité montrent que la proximité entre la coque et les microphones change la directivité des microphones. Au lieu d'avoir une caractéristique unidirectionnelle, ils ont plutôt une caractéristique omnidirectionnelle. Dû à cet effet, les microphones dirigés vers l'avant et l'arrière captent suffisamment les sons provenant des côtés. En plus, les microphones latéraux ($M2$ et $M5$) captent les sons provenant de l'avant et de l'arrière de la même manière, réduisant aussi l'information directionnelle. Ainsi dans la suite on ne prend plus en compte les signaux des microphones latéraux ($M2$ et $M5$). On filtre les signaux des microphones avec un HSF de 3 kHz dont le gain dépend de la direction. Pour les microphones positionnés à l'avant ($M1$ et $M6$) on applique un gain de 6 dB. Pour les microphones à l'arrière ($M3$ et $M4$) on applique un gain de -6 dB [76]. Pour obtenir le signal stéréo de sortie on additionne les 4 signaux, filtrés par côté.

Le **Prototype B** est équipé d'un microphone ambisonique d'ordre 1, monté au milieu de l'arc qui tient ensemble les coques. Avec le logiciel "VVMic" de *VVAudio*, on implémente l'encodage ambisonique et le décodage vers quatre haut-parleurs, placés virtuellement et rectangulairement dans l'espace. On

procède comme pour le Prototype A : filtrage avec un HSF de 3 kHz. Les gains pour les signaux des haut-parleurs placés virtuellement à l’avant, et à l’arrière, sont respectivement de 6 dB et -6 dB. Les 4 signaux filtrés sont additionnés par côté.

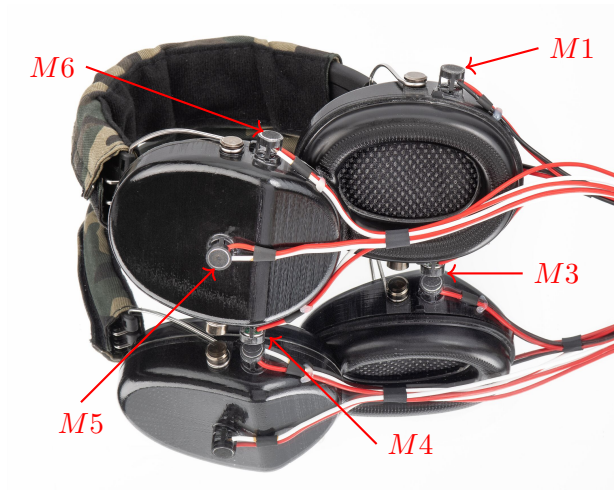
Les six microphones du **Prototype C** sont montés régulièrement dans un plan parallèle au plan horizontal sur un casque balistique. Les microphones sont du même type que pour le Prototype A. On filtre les signaux des microphones x_{M1} à x_{M6} avec l’HRTF \mathcal{H} de la tête artificielle KEMAR [140] en ne prenant en compte que les données ipsilatérales. Le signal stéréo de sortie est obtenu en suivant l’Équation (A.3).

$$\begin{aligned} L_C &= \frac{1}{2} \cdot \mathcal{H}(0^\circ) * x_{M1} + \mathcal{H}(60^\circ) * x_{M2} + \mathcal{H}(120^\circ) * x_{M3} + \frac{1}{2} \cdot \mathcal{H}(180^\circ) * x_{M4} \\ R_C &= \frac{1}{2} \cdot \mathcal{H}(0^\circ) * x_{M1} + \mathcal{H}(60^\circ) * x_{M6} + \mathcal{H}(120^\circ) * x_{M5} + \frac{1}{2} \cdot \mathcal{H}(180^\circ) * x_{M4} \end{aligned} \quad (\text{A.3})$$

Le **Prototype D** est inspiré de l’oreille externe humaine [132, 166]. On creuse une cavité qui ressemble à la conque dans l’extérieur des coques, en suivant les géométries données dans [147]. Au centre de la conque, on place un microphone qui ressemble au *Kingstate* KEIG4537TFL-N mais qui a une caractéristique omnidirectionnelle. De cette façon, les sons sont déjà filtrés par la géométrie de la coque avant qu’ils n’arrivent aux microphones, comme dans la situation où les sons passeraient du pavillon de l’oreille externe au tympan.

Avec les prototypes, une protection industrielle et la tête artificielle nue, des mesures dans la cabine semi-anechoïque ont été réalisées pour évaluer les indices spectraux et enregistrer des stimuli pour un dernier test de localisation. Les conditions d’écoute (HC) pour ces mesures sont expliquées dans la Figure A.10.

Pour toutes les conditions d’écoute, la différence des **indices spectraux** entre les directions médianes arrière et avant est proche de zéro pour des fréquences inférieures à 1 kHz, c.f. Figure A.11. Au-delà de 10 kHz la bande passante des microphones limite la transmission. Entre 1 kHz et 10 kHz les différences spectrales sont maximales. Les positions des maxima correspondent à la bande de Blauert de direction avant entre 1.86 kHz et 7.03 kHz [40]. Seul HC5 (Prototype D) ne dispose pas d’une différence spectrale comme HC0 à HC4. Du fait que le microphone soit positionné assez profondément dans la cavité, il reçoit des ondes qui sont diffractées au même bord de la cavité et donc ont les mêmes indices spectraux pour la majorité des positions sonores



(a) Prototype A avec les microphones $M1$ à $M6$. Le microphone $M2$ se trouve à l'extérieur de la coque gauche dans la direction opposée et n'est donc pas visible ni marqué.



(b) Prototype B avec le microphone ambisonique d'ordre 1. *AmbMic* dénomme les capsules arrangées tétraédriquement du microphone ambisonique.

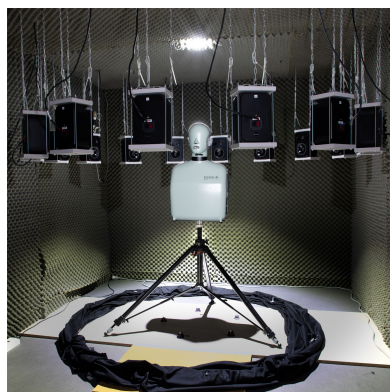


(c) Prototype C avec ses microphones. Les microphones $M3$ et $M4$ ne sont ni visibles ni indiqués sur la figure.



(d) Prototype D avec la cavité, imitant la conque, dans l'extérieur de la coque. L'intérieur de la coque droite est visible, donnant vue sur l'arrière du microphone $M2$.

FIGURE A.9 : Présentation des quatre prototypes qui ont été développés dans le contexte de l'étude. Les prototypes sont posés sur un miroir.



(a) HC0



(b) HC1



(c) HC2



(d) HC3



(e) HC4



(f) HC5

FIGURE A.10 : Figure A.10a : tête artificielle HATS33 montée dans la cabine semi-anechoïque, prête pour des mesures de HC0. Les mesures pour HC1 à HC5 sont réalisées de manière similaire en mettant la protection auditive ou le prototype sur la tête artificielle. Figures A.10b à A.10f : vues détaillées de la tête artificielle portant HPD P2 (Figure A.10b) et les prototypes A à D (Figures A.10c à A.10f).

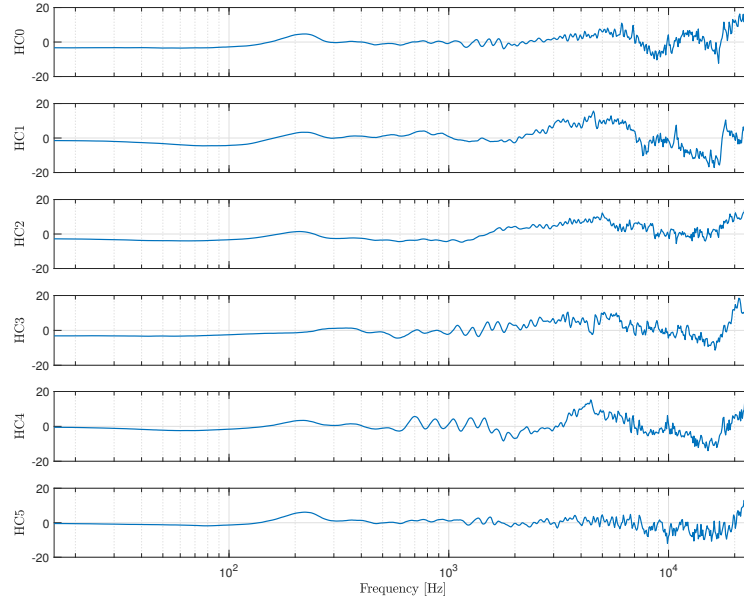


FIGURE A.11 : amplitude de la différence spectrale avant-arrière ΔH pour des positions médianes $\phi = 0^\circ$ et $\phi = 180^\circ$ pour les six conditions d'écoute HC0 to HC5. Les courbes sont centrées en zéro.

Avec un **test d'écoute**, on évalue la capacité des sujets à discriminer l'avant et l'arrière en leur demandant d'écouter des sons évoluant à l'intérieur d'un quadrant et d'indiquer le quadrant et le sens de rotation de chaque son. Les sons sont enregistrés dans la cabine anechoïque et traités pour chaque condition d'écoute. Les sons commencent à $\phi_t \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, tournent de 90° dans le sens des aiguilles d'une montre ou dans les sens contraire et sont diffusés par un casque audio. Après la diffusion d'un son, les sujets peuvent répondre directement ou choisir de ré-écouter le son qui vient d'être diffusé avant de répondre. Le test est géré automatiquement par un logiciel, développé en incluant une interface pour interagir avec les sujets. Le test est séparé en deux parties : dans la partie 1, on utilise une parole masculine, dans la partie 2 un bruit blanc pulsé. Chaque sujet participe aux deux parties dont l'ordre est décidé aléatoirement. D'abord on observe que le temps de réponse diminue avec l'expérience car les sujets s'habituent au test. En plus le type de son n'a pas d'influence sur la performance de localisation. Par contre, le son préféré par les sujets correspond dans 72.2 % au son de la deuxième partie. En moyenne on observe un taux de réponses correctes de 51.41 % et un taux de confusions avant-arrière de 40.83 %. Le taux de réponses correctes varient selon les conditions d'écoute entre 61.40 % (HC2), 54.69 % (HC0), 53.70 % (HC3), 50.41 % (HC4), 44.73 % (HC5) et 43.52 % (HC1).

Un test de Kruskal-Wallis montre que la différence entre les conditions d'écoute est significative sur le taux de réponses correctes (p -value $\ll 0.001$). Les confusions avant-arrière varient entre 30.96 % (HC2), 36.98 % (HC0), 37.32 % (HC3), 41.09 % (HC4), 48.50 % (HC5) et 49.48 % (HC1). D'après un test de Kruskal-Wallis, ces différences sont également significatives (p -value $\ll 0.001$). Avec tous les prototypes on observe une possibilité d'amélioration de la performance de localisation par rapport à l'HPD P2 (HC1). Sur l'ensemble des prototypes testés, c'est avec le Prototype A (HC2) qu'on obtient les meilleurs performances de localisation. Dans le cas où les sujets se trompent entre l'avant et l'arrière, ils répondent aussi dans 85.31 % des cas avec une confusion du sens de direction. Refléter le quadrant au plan frontal (équivalent à une confusion avant-arrière) mais garder le mouvement latéral-médian induit forcément une confusion du sens de rotation.

A.6 Conclusion

Dans cette étude, on présente le travail mené pour répondre à la question initiale : comment améliorer la localisation d'une source sonore en utilisant des protections auditives. Trois tests subjectifs ont été réalisés, accompagnés par des mesures objectives. Des protections disponibles sur le marché ont été évaluées et des prototypes ont été construits. On a pu faire ressortir les réponses suivantes par rapport à la question initiale :

- La performance de localisation dépend très fortement du type de protection : les bouchons sont plus performants que les casques. Dans le contexte de notre étude, les trois protections actives confèrent une meilleure performance que la protection passive.
- Dans le cas des bouchons, il faut arriver à une taille de bouchon suffisamment petite pour qu'il couvre le moins possible l'oreille externe. S'il peut être placé dans le conduit auditif, cela serait encore mieux.
- Dans le cas des casques, la performance de localisation peut être récupérée en induisant des indices spectraux par des filtres numériques.
- Des filtres numériques de bas ordre (inférieurs à l'ordre 10) donnent de meilleurs résultats de localisation que des HRTFs non-individuelles de têtes artificielles.

A.6. CONCLUSION

- Il est possible de construire des protections auditives de type casque qui permettent d'améliorer la performance de localisation en leur appliquant un design adéquate.

A.6. CONCLUSION

Appendix B

Ambisonic encoder test

Ambisonic recordings were used extensively in combination with non-individual filters for listening tests in Chapter 4. Therefore, it is of importance to investigate the localization performance of subjects listening to such ambisonic signals with their individual HRTF, i.e. natural hearing. The aim of this appendix is to assess the impact of different ambisonic signals on the localization performance.

Ambisonic B-format signals can be derived from measured A-format signals but also obtained by synthesis. During the measurements, the sound sources are physically placed around the microphone. The measurement setups require precise orientation as it highly influences the perceived location of the sound sources during reconstitution. Signals of indoor measurements comprise also the room response. As by the nature of measurements, uncertainties are inevitable, whereas synthesis provides exact, numerical control of the positions of the sound source. Additional effort is required to apply room characteristics such as reflections. This appendix examines the difference between ambisonic signals measured in a semi-anechoic chamber and synthesized ambisonic signals in terms of subjective localization performance and the signals themselves.

B.1 Stimuli

Two sets of stimuli are prepared each using a different method for obtaining the encoded B-format signals. The 1st, resp. 2nd set of stimuli is based on acoustic measurements, resp. synthesis. For the measurements of the 1st set, a 1st order ambisonic microphone (TetraMic by *Core Sound*) is placed in the center of the circular loudspeaker array, c.f. Section 1.4.2.2. The microphone is oriented upright,

facing with its front loudspeaker N° 1. Male speech¹ is presented with each loudspeaker subsequently and recorded with the ambisonic microphone. The recorded A-format signal, i.e., the output signals of the microphone capsules **Front-Left-Up**, **Front-Right-Down**, **Back-Left-Down**, and **Back-Right-Up**, captured the sound source located at $\phi_t = \{0^\circ, 22.5^\circ, \dots, 337.5^\circ\}$. The ambisonic A-format signals are equalized and encoded to ambisonic B-format following Equation (B.1) [209].

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} FLU \\ FRD \\ BLD \\ BRU \end{bmatrix} \quad (\text{B.1})$$

B-format signals comprise the sound pressure W and the sound pressure gradients X , Y , resp. Z along the X-, Y-, resp. Z-axis [209]. The subsequent decoding is done for a squared loudspeaker setup with loudspeakers located at $\phi_s = \{45^\circ, 135^\circ, 225^\circ, 315^\circ\}$. The VVMic tool from *VVAudio* is used for microphone equalization, encoding, and decoding [210].

The 2nd set of stimuli is based on synthesizing the ambisonic B-format signals. According to Benjamin et al. [207], B-format signals are basically obtained by trigonometric weighting of the mono source signal S with the horizontal, resp. vertical angle of incidence ϕ , resp. θ .

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \cos(\phi) \cos(\theta) \\ \sin(\phi) \cos(\theta) \\ \sin(\theta) \end{bmatrix} \cdot S \stackrel{\theta = 0}{=} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \cos(\phi) \\ \sin(\phi) \\ 0 \end{bmatrix} \cdot S \quad (\text{B.2})$$

Based on Equation (B.2), the ambisonic B-format signals are calculated for the source signal S (male speech, c.f. 1st set) at positions of $\phi = \phi_t = \{0^\circ, 22.5^\circ, \dots, 337.5^\circ\}$. The B-format signals are decoded for a squared loudspeaker setup, c.f. 1st set.

B.2 Experimental procedure

The test is designed as a forced-choice test where listeners are seated in the center of the circular loudspeaker array, c.f. localization test in Section 2.3. Each subject is listening to 32 stimuli (2 sets of 16 stimuli). Stimuli are presented in subgroups which are corresponding to the two sets, i.e., stimuli of different sets are not mixed. The intra-set order of the stimuli is random, as well as the order of the sets. Each stimulus is presented once without repetition. Even though the loudspeaker array

¹Harvard Sentences, List 19, Sentence 6: “Add the column and put the sum here.”. Duration: 3 s.

provides 16 loudspeakers, only those four loudspeakers are used for reproduction which correspond to the squared loudspeaker setup of the ambisonic decoding stage. Thus, the sound emitting loudspeakers are located at $\phi = \{45^\circ, 135^\circ, 225^\circ, 315^\circ\}$.

Upon arrival, participants are instructed to the test. During the test they are asked to indicate the number of the loudspeaker which is in the direction where they perceive the presented sound. They are delivered with a response sheet, c.f. Figure B.1, to write down the answers. Modifying previous answers is not allowed, even if the current sound makes them realize having falsely located preceding sounds. There is no time limit for responding. Further, they are told that the positioning of the head is very important and minor displacements will lead to false perceptions. They are instructed how to correctly position the head in the center point of the loudspeaker array, remaining in this position during listening, and reposition after having responded. The test supervisor triggered the computer to present the next sound after the subject verbally indicated having answered and replaced their head.

B.3 Results

25 listeners participated in the test, whereof 11 listened to the recorded signals first and 14 listened to the synthesized signals first. In total, 800 data sets (16 directions, 2 sets, 25 participants) were recorded.

B.3.1 Angular error

A Kolmogorov-Smirnov test proves that the angular error is normally distributed (p-value: 0.013), c.f. the distribution of the signed angular error as shown in Figure B.2a. It is obtained that 44.25 % of the sounds are localized correctly, i.e., $\epsilon = 0^\circ$, while 45.50 % of the sounds are localized at neighboring positions, i.e., $\epsilon = \pm 22.5^\circ$. For 89.75 % of the responses the unsigned angular error is $|\epsilon| \leq 22.5^\circ$, corresponding to the rate which was obtained by Joubaud [166]. They were using eight loudspeakers to generate 16 source positions and obtained in 95.38 % of all responses an unsigned angular error of $|\epsilon| \leq 22.5^\circ$ [166].

The trend of the unsigned angular error $|\epsilon|$ over the test progress, averaged over all 25 subjects, and its moving average over 5 samples are shown in Figure B.2b. $|\epsilon|$ varies along the test progress in the interval $[5.40^\circ; 20.70^\circ]$ with an average of $\mu = 15.24^\circ$ and a standard deviation of $\sigma = 3.73^\circ$.

B.3. RESULTS

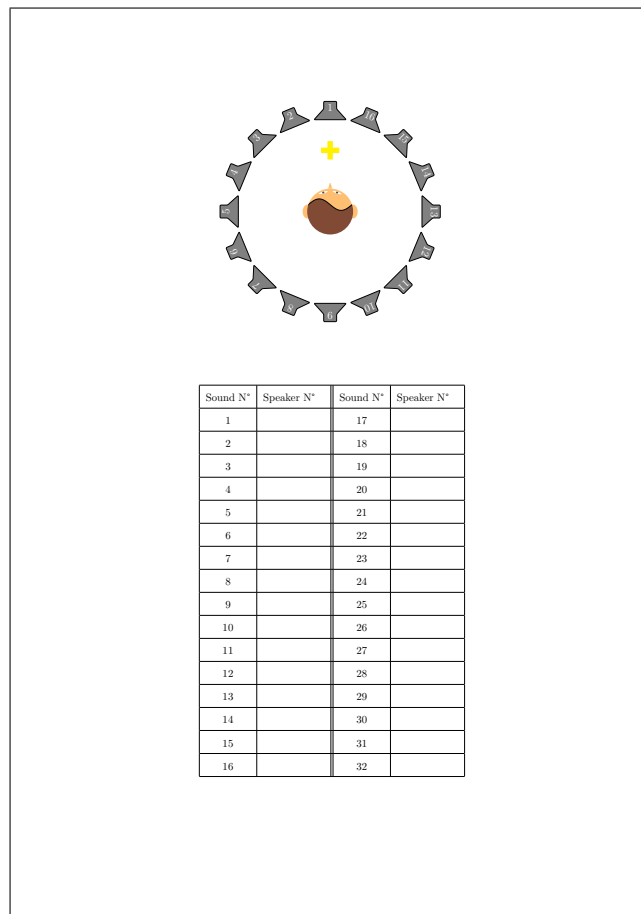


Figure B.1: Response sheet which is provided to subjects.

B.3. RESULTS

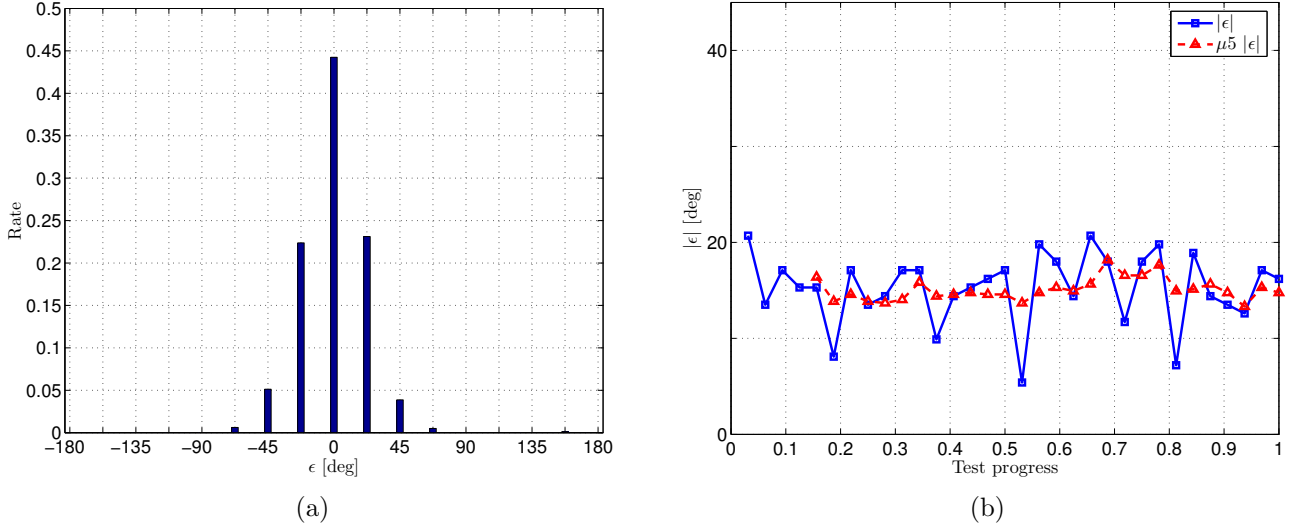


Figure B.2: Distribution of the signed angular error ϵ (Figure B.2a). Trend of the unsigned angular error $|\epsilon|$ and its moving average over 5 samples $\mu_5 |\epsilon|$ along the test progress (Figure B.2b).

$|\epsilon|$ does not noticeably differ between the first 16 sound samples, i.e., test progress $tp \leq 0.5$, and the last 16 sound samples, i.e., test progress $tp > 0.5$ ($\mu_{tp \leq 0.5} = 15.13^\circ$, resp. $\mu_{tp > 0.5} = 15.36^\circ$). The obtained standard deviation σ is slightly increased during the 2nd half of the test ($\sigma_{tp \leq 0.5} = 3.00^\circ$, resp. $\sigma_{tp > 0.5} = 4.45^\circ$). As more measurement-based stimuli are presented during the 2nd half, the increased value of σ seems to be related to a less precise localization performance which is caused by these measurement-based stimuli. $|\epsilon|$ and its moving average show both a constant trend along the test progress. It is not obtained that the unsigned angular error is dependent on the test progress.

B.3.2 Confusion Matrix

The confusion matrices in Figure B.3 visualize for both types of stimuli the distributions of the responded angle ϕ_r over each tested angle ϕ_t . A circle indicates that the corresponding pair of (ϕ_t, ϕ_r) occurs at least once in the obtained data set. The size and color of the circles are direct proportionally to the relative occurrence of the pairs of (ϕ_t, ϕ_r) . The main diagonal, which corresponds to perfect localization, and two anti-sub-diagonals, which correspond to idealized front-back errors, are traced with light gray lines. It is seen that pairs of (ϕ_t, ϕ_r) with a high occurrence are located either on or near the main diagonal. In the upper-left quadrant, i.e., $\phi_t < 180^\circ$ and $\phi_r > 180^\circ$, and the lower-right quadrant, i.e., $\phi_t > 180^\circ$ and $\phi_r < 180^\circ$, only little data are obtained. Data are mostly located in the lower-left quadrant, i.e., $\phi_t, \phi_r < 180^\circ$, and upper-right quadrant, i.e., $\phi_t, \phi_r > 180^\circ$. Thus, the

B.3. RESULTS

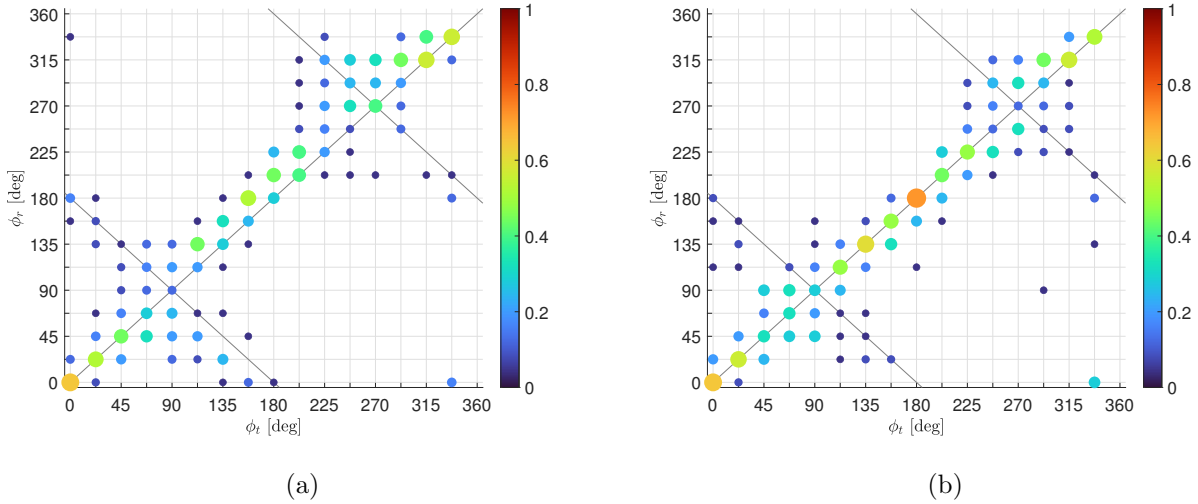


Figure B.3: Confusion matrices for the measurement-based (Figure B.3a) and synthesis-based (Figure B.3b) stimuli.

sound source localization shows an overall good performance and localization errors are mainly due to front-back confusions. Left-right and mixed confusions are obtained rarely. By visual inspection it is noticed that the distributions of ϕ_r vary along ϕ_t . We will pick up this fact in the next section. Further, differences are obtained between the two types of stimuli, in particular for left-back positions, i.e., $\phi_t \in [112.5^\circ; 180^\circ]$. Good localization performance is obtained with the synthesized stimuli while the measurement-based stimuli lead to a localization bias of 22.5° towards back positions.

B.3.3 Angular error per direction

The confusion matrices showed that the distributions of ϕ_r depend on ϕ_t . This requires investigating the angular error as a function of ϕ_t , too. The signed angular error ϵ over the tested angle ϕ_t is shown in Figure B.4, for both sets of stimuli. ϵ closest to 0° are obtained at $\phi_t = \{67^\circ, 112.5^\circ, 135^\circ, 202.5^\circ\}$ (measured stimuli), resp. $\phi_t = \{247.5^\circ\}$ (synthesized stimuli). Extreme values of the angular error are obtained at lateral positions of $\phi_t = 90^\circ$ (measurement: $\epsilon = -32.4^\circ$, synthesis: $\epsilon = -25.2^\circ$) and $\phi_t = 270^\circ$ (measurement: $\epsilon = 22.5^\circ$, synthesis: $\epsilon = 25.2^\circ$), c.f. Figure B.4. Humans generally show poor localization performance at lateral positions, c.f. an increased spread of ϕ_r at $(\phi_t) = \{90^\circ, 270^\circ\}$ in Section 2.3.3. Benjamin et al. identified squared loudspeaker arrangements to be least preferable than rectangular or hexagonal setups regarding the localization performance at lateral positions [207]. In the interval $\phi_t \in [112.5^\circ; 202.5^\circ]$ the angular errors of the measured and

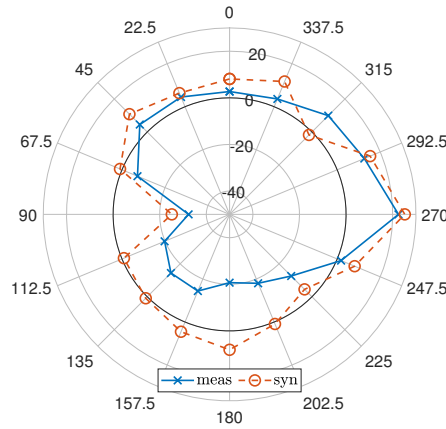


Figure B.4: Polar plot of the signed angular error ϵ (coded by the polar radius in the graph) for the 16 tested angles ϕ_t (coded by the polar angle in the graph). The polar radius ranges the interval $[-50; 30^\circ]$ and $\epsilon = 0^\circ$ is highlighted by a black circle.

synthesized stimuli highly diverge. At these directions the offset between the curves in Figure B.4 range between 15.3° ($\phi_t = 135^\circ$) and 28.8° ($\phi_t = 180^\circ$).

For sounds which are positioned in the frontal hemisphere ($\phi_t \in [0^\circ; 90^\circ[\cup]270^\circ; 360^\circ]$) a negative, signed angular error represents a clockwise shift between ϕ_t and ϕ_r . For sounds which are positioned in the back hemisphere ($\phi_t \in]90^\circ; 270^\circ[$) negative, signed angular errors correspond to a counterclockwise shift between ϕ_t and ϕ_r . Hence, measured, resp. synthesized stimuli are resulting in 81.25%, resp. 50% to a counterclockwise shift of the responded angle ϕ_r . The localization bias of 22.5° for the measurement-based stimuli, as identified in the confusion matrix, is hereby proven and most important for back positions, denoting a counterclockwise shift.

B.3.4 Localization errors

A left-right, front-back, resp. mixed confusion rate of 0.43%, 16%, resp. 0.17% was obtained over the entire data. The low rates of left-right and mixed errors allow to conclude that, first, the provided binaural cues of the ILD and ITD were consistent during the localization test, and second, subjects correctly placed their head on the Y-axis.

In the following only the front-back confusions are considered which equal 20%, resp. 11% for the measured, resp. synthesized signals. The type of stimuli shows a significant effect on the front-back confusions (p-value = 0.0095), see also non-overlapping notches in Figure B.6. The increased

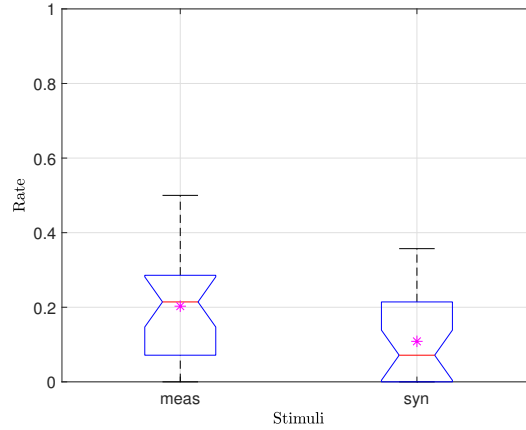


Figure B.5

Figure B.6: Box plot with mean values (asterisks) of the front-back confusion rates for the measurement (meas) and synthesis (syn) based stimuli.

front-back confusion rate during the measured stimuli is subjected to be caused by a microphone positioning not matching perfectly with the center point of the loudspeaker array. In addition, the room response is an inevitable part of the measured signals. The acoustical properties of the semi-anechoic chamber interact with the placement of the loudspeaker array leading to reflections which depend on the direction of incidence ϕ_t .

20 % (measured stimuli) and 16 % (synthesized stimuli) of the median-frontal directions ($\phi_t = 0^\circ$) are localized in the back hemisphere ($\phi_r \in]90^\circ; 270^\circ[$). In contrast only 4 % (measured stimuli) and 0 % (synthesized stimuli) of median-back directions ($\phi_t = 180^\circ$) are localized in the frontal hemisphere ($\phi_r \in [0^\circ; 90^\circ[\cup]270^\circ; 360^\circ]$). Subjects are suspected having placed their heads slightly misplaced on the X-axis, towards the back. This issue should be investigated in further studies in combination with a head tracking system, such as used during the introductory localization test in Section 2.3.

B.4 Encoded signals

The measured and synthesized ambisonic B-format signals are visualized in Figure B.7. The gain of W_{meas} , resp. Z_{meas} ranges the interval $[0.58; 0.81]$ ($\mu = 0.68$), resp. $[-0.39; -0.16]$ ($\mu = -0.25$). The peak values of X_{meas} and Y_{meas} show both sinusoidal characteristics and X_{meas} an additional phase shift of 90° . Minima of X_{meas} , resp. Y_{meas} are located at 180° , resp. 270° and maxima are

located at 0° , resp. 90° . None of both signals show zero gain at any of the 16 discrete values of ϕ_t . Gains closest to zero are identified in X_{meas} at $\phi_t = \{67.5^\circ; 247.5^\circ\}$ and in Y_{meas} at $\phi_t = \{0^\circ; 157.5^\circ\}$.

According to Equation (B.2), the synthesized signals W_{syn} , resp. Z_{syn} show constant gains, c.f. Figure B.7b. The gains of X_{syn} , resp. Y_{syn} show perfect co-sinusoidal, resp. sinusoidal characteristics, where extreme values are located at $\phi = \{0^\circ; 180^\circ\}$ (X_{syn}) and $\phi = \{90^\circ; 270^\circ\}$ (Y_{syn}), and zero gains occur at $\phi = \{90^\circ; 270^\circ\}$ (X_{syn}) and $\phi = \{0^\circ; 180^\circ\}$ (Y_{syn}).

The average gain of W_{meas} ($\mu = 0.68$) is close to the averaged gain of W_{syn} ($\mu = 0.71$), showing good accordance and a correctly measured W_{meas} signal. The variations of the gain in W_{meas} are due to reflections in the semi-anechoic chamber, such as at the loudspeaker array. The gains in the up-down gradient Z_{meas} are all non-zero and negative. This indicates that there is a vertical component (c.f. non-zero) in the sound field which originates from below the microphone (c.f. negative). Reflections at the floor of the semi-anechoic chamber and the microphone stand are causing these gradients. Further, the position of the microphone seems to introduce these gradients, too. The mounting height of the loudspeaker array corresponds to the geometric center points of the individual loudspeakers. The loudspeakers consist of two transducers and a crossover frequency of 4.2 kHz [141], hence they are not point sources. The housing of the loudspeaker holds in its upper, resp. lower part the tweeter, resp. subwoofer. Therefore, high frequencies are emitted above while low frequency are emitted below the ambisonic microphone, causing the non-zero up-down gradient Z_{meas} .

The extreme values in X_{meas} , resp. Y_{meas} are correctly positioned at $\phi_t = \{0^\circ; 180^\circ\}$, resp. $\phi_t = \{90^\circ; 270^\circ\}$, in contrast they do not contain zero gains as they are supposed to do. The gains, which are the closest to zero, are clockwise shifted by 22.5° . During playback this shift causes the acoustic scene being counterclockwise rotated by 22.5° . The consequences are subjectively, falsely perceived directions of incidence, c.f. Figures B.3a and B.4. Reverberations and reflections at the floor of the semi-anechoic chamber and the relative placement of the loudspeaker array in the semi-anechoic chamber seem to interact with the ambisonic measurement. In addition, corrupted directivity patterns of the microphone capsules may privilege the microphone's sensitivity towards certain directions of incidence. Further the perfect positioning and orientation of the ambisonic microphone during measurement is of extreme importance and is highly error-prone [209].

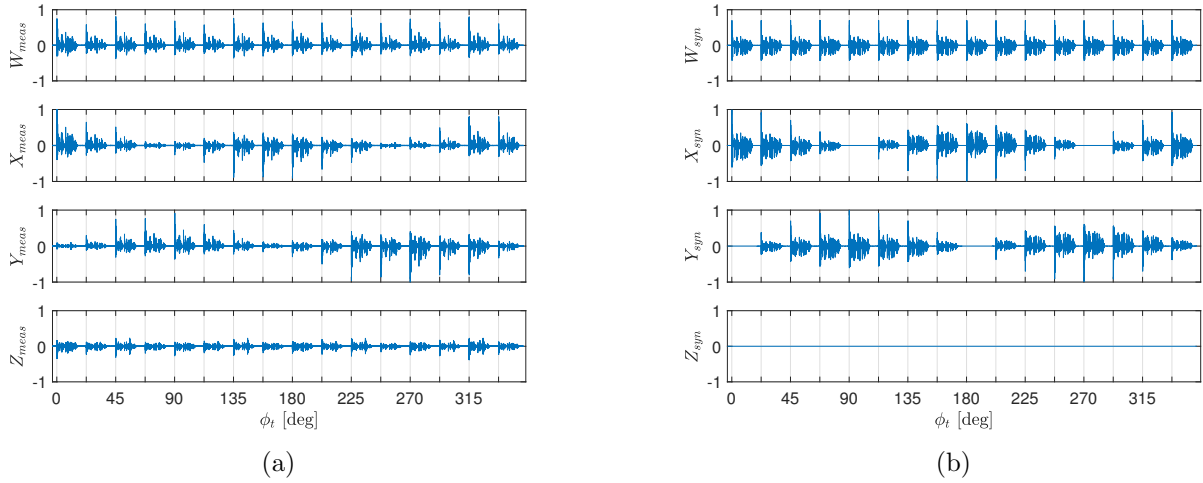


Figure B.7: Ambisonic B-Format signals for 16 discrete, horizontal angles ϕ_t . The signals are obtained by encoding the measured A-Format signals (Figure B.7a) and by synthesis (Figure B.7b).

B.5 Conclusion

The localization performance is significantly different between the measured and synthesized signals. Good localization performance was obtained using synthesized B-format signals, while reduced localization performance was obtained with ambisonic measured B-format signals. A localization bias of 22.5° for back sound positions and an increased front-back error rate are identified using the measurement-based signals. This difference is identified to be caused by the encoded B-format signals, mostly influenced by the two horizontal signal components X_{meas} and Y_{meas} . Small misalignment of the ambisonic microphone impacts the subjective localization of sound sources in the reconstructed sound field. In Chapter 4, the same measurement, resp. recording, setup to acquire the ambisonic signals was used. Though, the localization performance obtained in Chapter 4 is not only influenced by the deployed rendering filters, but also by misalignment of the setup which has large impact on the localization performance.

Regardless of measuring or synthesizing the B-format signals, small misalignment of the head instantly increases the confusion rates. Measuring, resp. recording the sound field introduces the additional critical aspect of the microphone positioning. The listener positioning during reconstruction must be precisely matched to either the positioning of the microphone during measurement or the reference point during synthesis. The positioning is less critical on the front-back confusion rates

B.5. CONCLUSION

when using discrete sound sources (6.8%), c.f. Chapter 2, than when using ambisonic signals (11%).

Nevertheless, synthesizing B-format signals shows advantages over encoding measured A-format signals. Contrary to measurements, synthesis is not facing the issues of microphone positioning, loudspeaker properties, and possibly disturbing room reflections. Therefore, synthesis results in accurate signals and consequently better localization of the sound direction.

B.5. CONCLUSION

Appendix C

List of acronyms

ANC Active Noise Control.

ANOVA Analysis of variance.

ANR Active Noise Reduction.

BEM Boundary Element Method.

BRIR Binaural room impulse response.

CGS Conjugate-Gradient Solver.

Cnam Conservatoire national des arts et métiers.

CoH center of head.

DRIR Directional room impulse response.

DTF Directional Transfer Function.

FEM Finite Element Method.

FIR Finite Impulse Response.

GUI Graphical User Interface.

HINT Hearing in Noise Test.

HPD Hearing Protection Device.

HRIR Head-Related impulse response.

HRTF Head-Related Transfer Function.

HSF High-Shelf Filter.

IIR Infinite Impulse Response.

ILD Interaural Level Difference.

ISL French-German Research Institute of Saint-Louis.

ITD Interaural Time Difference.

LDL Loudness Discomfort Level.

LMSSC Structural Mechanics and Coupled Systems Laboratory.

LSF Low-Shelf Filter.

MLFMM Multi-Level Fast Multipole Method.

MRT Modified Rhyme Test.

PC Personal Computer.

PCC Pearson's correlation coefficient.

PRTF Pinna-Related Transfer Function.

RMS Root Mean Square.

RMSE Root Mean Square Error.

SNR Signal-to-Noise Ratio.

SOFA Spatially Oriented Format for Acoustics.

SPL Sound Pressure Level.

List of acronyms

TFOE Transfer Function of the Open Ear.

TOA Time Of Arrival.

VAE Virtual Acoustic Environment.

Appendix D

List of symbols

L_{phon} Loudness in phon.

ϕ Azimuth angle in spherical coordinates.

$L_{eq,t}$ Equivalent sound level for time interval t .

$L_{ex,8h}$ Daily (8 h) exposure sound level.

θ Elevation angle in spherical coordinates.

Abstract : Many professionals, e.g. mining and construction workers, ground crews or soldiers are exposed to impulsive or constant high level noise. In order to prevent hearing loss, they depend on hearing protection devices (HPDs). On the contrary, HPDs interfere with situational awareness and sound source localization. This contradiction makes users pondering between hearing loss and situational awareness. Often the latter dominates over the former. This work aims to bring in line hearing protection and situational awareness. A virtual acoustic environment (VAE) with 16 circularly, horizontally arranged loudspeakers is set up. Localization performance with commercially available HPDs, including active and passive earplugs and earmuffs, is assessed in the VAE with 40 subjects. Earplugs with small geometries show better results than large-sized earmuffs. These results coincide with the study on modifications of the Head Related Transfer Function (HRTF) caused by HPDs. Earplugs preserve many individual spectral cues, while earmuffs cancel out most of these cues. We compare methods of combining a simulated, generic HRTF with the simulated, individual Pinna Related Transfer Function. An analytic model of HRTFs, controlled by the azimuth angle, is developed. Respecting the limitations of embedded systems, regarding energy supply and computational power, 14 digital filters are defined. A headphone based listening test is conducted to rate these filters regarding subjective front-back discrimination performance, resulting in better performance with low order filters than with high order filters. We present 4 designs of advanced HPDs which are aimed to improve the sound localization performance. Prototypes are manufactured and evaluated in a subjective listening test with 36 participants, showing that it is possible to improve sound localization of a commercially available active HPD.

Keywords : Binaural filter, Front-back confusion, Hearing protection device, HRTF measurement, HRTF modifications, HRTF simulation, Localization performance, Non-individual HRTF, Subjective listening test.

Résumé : De nombreux professionnels sont exposés à des bruits impulsifs ou constants, de très forts niveaux. Pour se prémunir d'une perte auditive, ils portent des protecteurs auditifs. Il s'ensuit une réduction de la localisation des sources sonores, cependant celle-ci est importante, notamment pour des raisons de sécurité. L'objectif de cette étude est de concevoir des systèmes qui concilient la protection auditive, tout en gardant la perception d'espace. Un environnement acoustique virtuel (EAV) de 16 haut-parleurs disposés circulairement et horizontalement est mis en place pour tester les systèmes de protection. La performance de localisation avec des protections acoustiques de types actives et passives, bouchons et casques, disponibles sur le marché, est évaluée avec 40 sujets. Ce test a montré que les bouchons sont à préférer aux casques, selon ce critère de conservation des capacités de localisation. Ces résultats sont corrélés avec les modifications des Head Related Transfer Functions (HRTFs) introduites par les protecteurs auditifs. On montre que les bouchons conservent mieux que les casques les indices spectraux individuels. On compare 2 méthodes pour combiner une HRTF générique et pré-simulée, avec la fonction de transfert individuelle relative à la coque simulée. Un modèle analytique des HRTFs, contrôlé par l'angle d'azimut, est développé. Avec les contraintes imposées, en termes de ressources d'énergie et de puissance de calcul sur des systèmes embarqués, 14 filtres sont définis. Un test d'écoute a permis d'évaluer ces filtres, concernant la discrimination des sons émis devant et derrière l'auditeur, et les filtres d'ordre faible montrent de meilleurs résultats que ceux d'ordre élevé. On propose 4 approches pour des protecteurs auditifs avancés de type casque, qui ont comme but d'améliorer la localisation des sons. Ces prototypes ont été assemblés, et évalués grâce à un test d'écoute avec 36 participants qui montre qu'il est possible d'améliorer la performance de localisation d'un protecteur auditif du commerce.

Mots clés : Confusion avant-arrière, Filtre binaural, HRTF non-individualisée, Mesure d'HRTF, Modification d'HRTF, Performance de localisation, Protecteur auditif, Simulation d'HRTF, Test subjective d'écoute.