Diss. ETH No. 20206

MEASURING, PREDICTING AND IMPROVING THE PERCEPTION OF SPACE WITH BILATERAL HEARING INSTRUMENTS

A dissertation submitted to the
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
MARTIN FELIX MÜLLER
Dipl.-Ing EPFL
born September 10, 1982
citizen of Mettauertal, AG

accepted on the recommendation of
Prof. Dr. Peter Bösiger, examiner
Prof. Dr. Norbert Dillier, co-examiner
Dr. Stefan Launer, co-examiner

2012

To Gabriela and Melissa, my honeybees...

Abstract

A number of studies have shown that bilateral hearing aids (i.e. two independent hearing devices) distort the perception of spatial characteristics of the acoustic space. The correct position of sound sources and their perceived width are modified by the signal processing in the devices. Furthermore, it was reported that sound sources are often perceived in the head of the hearing aid users rather than in the external auditory space. This observation was indeed confirmed by a pilot experiment, in which the listeners had to report the spatial perception of various auditory objects. The signals were processed by different hearing aid algorithms. The results showed that the hearing aid algorithms did alter the way in which the sounds were perceived by the listeners.

Earlier experiments that evaluated hearing prostheses were carried out in mostly simplified and artificial test environments which are not representative of situations where hearing aids are generally used. For a proper evaluation of the devices new methods need to be developed that are able to reproduce any acoustic setting in a perceptually plausible manner. To address this issue, a system for virtual acoustics was developed. This system combines Head-Related Transfer Functions (HRTFs), room simulations and the accurate reproduction of head movements for generating virtual acoustical spaces. In a first stage of the thesis, the system was evaluated with a number of experiments. In various conditions, the fidelity of sound reproduction was such, that the listeners had difficulties to distinguish between real and simulated playback for speech and noise stimuli.

A selection of state-of-the art hearing aid algorithms was combined with the virtual acoustics system and evaluated. Localization and distance perception experiments were carried out in realistic and complex simulated environments. The results do confirm previous studies carried out in much simpler settings and show that the algorithms have a significant impact on sound localization. In particular, the number of front-back confusions was at chance level for two of the algorithms tested. Additionally, the influence of head movements on sound localization for bilateral Cochlear Implants (CIs) users was investigated. The study demonstrated that CI users can use head movements to reduce the number of front-back confusions provided the signal is long enough. However, they are not able to use head movements to increase the angular acuity of their localization performance.

While the perceptual experiments cited above allow the evaluation of spatial auditory perception with auditory prostheses in realistic settings, they are very time consuming. To reduce the testing time a model of the binaural auditory system was implemented and combined with a statistical classifier to predict the spatial auditory perception for arbitrary binaural signals. The model consists of a peripherical stage and a binaural processor [Breebaart *et al.* 2001]. After a frequency band decomposition of the input signals, the binaural processor estimates

the best Interaural Time and Level Difference (ITD and ILD) at each time frame. The output of the binaural model is classified by a set of random forests trained to 710 positions in space at a given center frequency. The classifier results across all positions are displayed in perceptual maps. These maps characterize the spatial perception of a human listener for a binaural signal. Predictions of source width, sound localization and the probability of front-back confusions are related to the energy distribution of the perceptual maps.

The signals used in the perceptual experiments were run through the binaural model. The resulting perceptual maps showed good accordance with the outcome of the subjective experiments. In particular, the front-back and localization uncertainties corresponded to the feedback comments of the test subjects. The effect of reverberation and interaural coherence on spatial auditory perception were investigated as well with the binaural model. The findings show that reverberation and the microphone positions have an effect on source width and the rate of front-back confusions. In the conditions tested, the interaural coherence however affected the width of the sources only.

Finally, a new binaural hearing aid algorithm is introduced and analyzed with the perceptual model in the last part of the thesis. The algorithms combines a binaural Multichannel Wiener Filter (MWF) with a dereveberation algorithm. Both algorithms were designed to explicitly consider the interaural cues. The performance of the algorithm was evaluated using the binaural Speech Intelligibility Index. An improvement in the Speech Reception Threshold of up to 5 dB was found for different room conditions. The algorithm was evaluated with the binaural model and it was found that the localization of the target sound source was preserved.

Zusammenfassung

Zahlreiche Studien haben gezeigt, dass die auditorisch-räumliche Wahrnehmung durch bilaterale Hörgeräte zerstört wird. Grund dafür ist die Signalverarbeitung in den Geräten. Sie ändert die natürlichen interauralen Zeit- und Pegeldifferenzen (ITDs und ILDs). Die wahrgenommene Position und Breite von Schallquellen wird dadurch modifiziert. Zudem lokalisieren Hörgeräteträger die Schallquellen häufig im Kopf anstatt in der äusseren akustischen Umgebung. Frühere in der Literatur beschriebenen Studien haben Hörsysteme oft nur unter sehr einfachen und künstlichen Bedingungen getestet, welche nicht repräsentativ sind für Alltagsumgebungen. Um Hörgeräte richtig evaluieren zu können, müssen neue Testverfahren entwickelt werden, in welchen die komplexen akustischen Verhältnisse realistisch wiedergeben werden.

In dieser Arbeit wurde ein System für virtuelle Akustik entwickelt, das die Reproduktion von beliebigen akustischen Szenen ermöglicht. Das System kombiniert individuelle kopfbezogene Übertragungsfunktionen (HRTFs), Raumsimulationen und die präzise Wiedergabe von Kopfbewegungen. Im ersten Teil dieser Arbeit wurde eine Reihe von Wahrnehmungsexperimenten durchgeführt, um das System zu evaluieren. In gewissene Bedingungen war die Genauigkeit des Systems so hoch, dass die Testpersonen Schwierigkeiten hatten, virtuelle und reale Quellen zu unterscheiden. Das System für virtuelle Akustik wurde dann benutzt, um eine Reihe von aktuellen Hörgerätealgorithmen unter realistischen akustischen Bedingungen zu testen. Subjektive Lokalisations- und Distanzwahrnehmungsexperimente wurden durchgeführt. Die Resultate bestätigen frühere Ergebnisse und zeigen, dass Hörgerätealgorithmen tatsächlich einen signifikanten Effekt auf die räumliche Wahrnehmung haben. Insbesondere lagen die vorne-hinten- Verwechslungen im Zufallsbereich für zwei von vier getesteten Algorithmen.

Das System wurde ebenfalls dazu verwendet, die Schalllokalisationsfähigkeiten von bilateralen Cochlea-Implantat-Trägern (CIs) in realistischen Bedingungen zu untersuchen. Aus der Literatur ist bekannt, dass Kopfbewegungen Schallquellen von vorne und hinten zu trennen helfen. Es war aber bisher unklar, ob die CI-Träger diese Kopfbewegungen effektiv ausnützen können. Ein Lokalisationsexperiment wurde durchgeführt, in welchem die Kopfbewegungen der CI-Träger aufgenommen und analysiert wurden. Die Ergebnisse zeigten, dass die CI-Träger diese Bewegungen ausnützen können, um die vorne-hinten-Verwechslungen zu reduzieren, wenn das Signal lang genug ist. Die Bewegungen hatten aber keinen Einfluss auf die azimutale Auflösung der Probanden.

Die räumliche Qualität von Hörgeräten lässt sich in realistischen Verhältnissen mit Wahrnehmungsversuchen gut abschätzen kann, allerdings mit erheblichen Zeitaufwand. Um die Messzeit zu reduzieren, wurde der Binaural Auditory System Simulator (BASSIM) ent-

wickelt. Der BASSIM kombiniert ein Modell des binauralen Hörsystems mit einem statistischen Klassifikator. Ziel des BASSIMS ist, voraussagen zu können, wie binaurale Signale von Normalhörenden räumlich empfunden werden. Das Modell basiert auf der Arbeit von Breebaart et al. (2001) und besteht aus einem peripheren Modell und einem binauralen Prozessor. Das periphere Modell bildet die Funktionen des äusseren, mittleren und inneren Ohres nach. Der binaurale Prozessor besteht aus Modulen, die die menschliche binaurale Verarbeitung simulieren. Das periphere Modell teilt die Eingangssignale in verschiedene Frequenzbänder auf. Danach findet der binaurale Prozessor für jedes Frequenzband und für jedes Zeitfenster die beste Kombination von interauralen Zeit- und Pegeldifferenzen, welche zum statistischen Klassifikator geschickt wird. Die Signale werden für jedes Frequenzband und für jedes Zeitfenster in eine von 710 trainierten Positionen klassifiziert (Elevationen -45^{o} bis 90^{o}). Die Klassifizierungsergebnisse können danach über alle Frequenzbänder, Zeitfenster und Positionen kombiniert werden und in so genannten perzeptuellen Darstellungen abgebildet werden. Diese Darstellungen zeigen, wo und wie eine bestimmte Quelle von einem menschlichen Zuhörer wahrgenommen wird.

Der BASSIM wurde dazu verwendet, die Signale des Lokalisationsexperiments zu analysieren. Die Klassifikationsergebnisse wurden mit den Resultaten des Experiments verglichen. Die perzeptuellen Darstellungen bestätigten die experimentellen Ergebnisse. Insbesondere stimmten die abgebildeten vorne-hinten- und Positionsunsicherheiten mit den Kommentaren der Probanden überein. Der BASSIM wurde mit verschiedenen zusaätzlichen Szenarien getestet. Der Effekt des Nachhalls und der Interauralen Kohärenz (IC) auf die perzeptuellen Darstellungen wurde untersucht. Die Ergebnisse zeigten, dass Nachhall und die Mikrofonposition einen Einfluss auf die vorne-hinten-Unsicherheit haben. Die interaurale Kohärenz hat nur einen Einfluss auf die wahrgenommene Breite der Schallquellen.

Am Ende dieser Arbeit wird ein neuer binauraler Algorithmus eingeführt. Der Algorithmus kombiniert ein binaurales mehrkanaliges Wiener Filter (MWF) mit einem Enthallungsalgorithmus. Beide Algorithmen wurden unter der Bedingung entwickelt, die interaurale Information explizit zu betrachten. Der Algorithmus wurde mit dem binauralen Speech Intelligibility Index (SII) und BASSIM evaluiert. Der SII zeigt eine 5 dB SRT (Speech Reception Threshold) Verbesserung für den Algorithmus in verschiedenen akustischen Umgebungen. Die Evaluationsergebnisse mit BASSIM zeigen, dass die perzeptive Lokalisation der Schallquelle erhalten blieb. Weitere Experimente sind allerdings noch nötig, um den Algorithmus unter realistischen Bedingungen zu evaluieren.

Acknowledgements

First of all, I would like to thank Prof. Peter Bösiger, Prof. Norbert Dillier and Dr. Stefan Launer without whom this work would not have been possible. To Norbert, I am especially thankful for his constant help, support and advice during the many years this project lasted.

During this work, I spent most of my time in a rather dull place full of wires and loudspeakers called "room7" where the listening experiments were carried out. Fortunately, I could drag Andrea into running most of those experiments for me. I am grateful to her for her help and fun way of working. She introduced me to the fascinating worlds of yoga and aromatherapy.

Apart from "room 7", I had the pleasure to share my office with Steven. I would like to thank him for the very interesting work and non-work related discussions. Also, he comforted me in my moments of doubts, restoring faith in our project. During the last months of the thesis, our solar-physicist René replaced him as my office-mate. I appreciated very much the few weeks we spent together in the office. This made leaving even more difficult.

I would also like to thank the Phonak crew: Juliane, Markus and Schelle. I especially appreciated the very interesting and always kind discussions we shared. They were always ready to offer their help whenever it was needed. Our bi-weekly brainstorming sessions provided the input I needed to push the project forward. Thank you for that!

To Katrin, that is just starting her PhD journey, I wish the best for the coming years! I am also grateful to her for the good and fun moments at the ORL-clinic and abroad. I would like to thank Tobias as well for his work on the real-time reproduction of head movements. He brought new ideas and new possibilities that have been partly explored in this thesis. This would not have been possible without him. Michael and Wai-Kong were further member of the LEA-research group. They accompanied my doctorate since the beginning. With their sense of humor, they made sure that the LEA was a nice place to be.

A special dedication to the people of the LEA-Kaffeegruppe I have not mentioned yet: Anita, Anja, Franzi, Lilian, Markus, Silva, Stefan and the ones I forgot. The nice atmosphere you created and the cakes you baked were responsible for the couple of extra-kilos I developed through the years.

Gaby, my love. I take this opportunity to thank you for staying at my side through all these years and cheering me up when morale was low. Because of you, this years produced some unexpected fruits and unique joys.

To my dear parents, you always believed in me and supported me. Part of this work is yours...

Contents

List of Acronyms xiii

1. Introduction 1
1.1. Objectives . . . 3
1.2. Thesis outline . . . 3

2. Binaural hearing 7
2.1. The peripherical Human Auditory System . . . 7
2.2. Binaural Localization . . . 10
2.2.1. Discrimination of Interaural Cues . . . 12
2.2.1.1. ITD and ILD threshold for pure tones . . . 12
2.2.1.2. Detection of high-frequency ITDs in SAM tones . . . 14
2.2.1.3. Cue discrimination experiments using other stimuli . . . 14
2.2.1.4. Band-importance of the interaural cues for sound localization 15
2.2.2. Binaural cues in a reverberant environment . . . 17
2.2.2.1. Precedence effect . . . 18
2.3. Auditory Source Width and Spaciousness . . . 19
2.3.1. Interaural cues and measures of spaciousness . . . 19
2.4. Sound internalization . . . 21
2.5. Models of the binaural auditory system . . . 22
2.5.1. Cross-correlation models . . . 22
2.5.2. Equalization-Cancelation models . . . 24
2.5.3. Using binaural models in complex listening situations . . . 25
2.6. Summary . . . 25

I. Test methodology 27

3. System for virtual acoustics 29
3.1. Introduction . . . 29
3.2. Virtual representation . . . 29
3.2.1. HRTF measurement and calibration procedure . . . 30
3.2.2. Measurement and calibration of the speaker-microphone system . . . 30
3.2.3. Calibration of the BTE system . . . 32
3.2.4. HRTF interpolation . . . 33

3.2.4.1. Room modeling and simulation . . . 34
3.3. Head-tracking and dynamic scene rendering . . . 35
3.3.1. Generating moving sources . . . 36
3.3.2. Updating head movements . . . 37
3.4. Perceptual evaluation . . . 39
3.4.1. Stimuli . . . 40
3.4.2. Procedure . . . 40
3.4.3. Results . . . 41
3.4.3.1. Externalization . . . 41
3.4.3.2. Stability . . . 42
3.4.3.3. Classification of the Sound Source . . . 43
3.4.3.4. Position dependency . . . 44
3.5. Conclusion . . . 45

II. Perception of the auditory space with bilateral hearing instruments 47

4. Localization with bilateral hearing aids 49
4.1. Introduction . . . 49
4.2. Method . . . 51
4.2.1. Reference conditions . . . 51
4.2.2. Hearing aid algorithms . . . 53
4.2.3. Virtual sound reproduction . . . 54
4.2.3.1. HRTFs measurements . . . 54
4.2.3.2. BTE HRTFs interpolation . . . 54
4.2.3.3. HRTF and BRTF calibration . . . 55
4.2.3.4. Room modeling and simulation . . . 55
4.2.4. Test procedure . . . 56
4.2.5. Test subjects . . . 56
4.2.6. Data analysis . . . 56
4.3. Results . . . 57
4.3.1. Experiment I: Evaluation of the virtual acoustics system . . . 57
4.3.2. Experiment II: Evaluation of BTEs algorithms . . . 62
4.3.3. Evaluation of hearing impaired subjects . . . 65
4.4. Discussion . . . 68
4.4.1. Sound localization in noise . . . 68
4.4.2. Localization of virtual sound sources . . . 69
4.4.3. Hearing aid localization . . . 70
4.5. Conclusion . . . 71

5. Head movements and localization with bilateral Cochlear Implants 73
5.1. Introduction . . . 73

5.2. Methods . . . 74
5.2.1. Test subjects . . . 75
5.2.2. Data analysis . . . 76
5.2.2.1. Analysis of localization performance . . . 76
5.2.2.2. Analysis of head trajectories . . . 77
5.3. Results . . . 78
5.3.1. Analysis of localization performance . . . 78
5.3.2. Analysis of head trajectories . . . 79
5.4. Discussion . . . 82
5.5. Conclusion . . . 84

6. Distance perception with bilateral hearing aids 87
6.1. Introduction . . . 87
6.2. Methods . . . 89
6.3. Results . . . 91
6.4. Discussion and Conclusion . . . 93

III. Predicting and improving algorithm performance 95

7. Predicting spatial perception 97
7.1. Binaural Auditory System Simulator . . . 97
7.1.1. BASSIM implementation . . . 97
7.1.1.1. Peripheral model . . . 97
7.1.1.2. Binaural processor . . . 99
7.1.1.3. Statistical classifier . . . 101
7.2. Prediction of spatial perception . . . 103
7.2.1. Localization prediction . . . 104
7.2.1.1. Influence of hearing aid algorithms on the perceptual maps . 106
7.2.2. Source width . . . 108
7.3. Conclusion . . . 109

8. Hearing aid algorithm for improved spatial perception 113
8.1. Introduction . . . 113
8.2. Algorithm . . . 114
8.2.1. Binaural noise reduction :Multi-channel Wiener filter . . . 114
8.2.2. Binaural dereveberation . . . 117
8.3. Implementation . . . 119
8.3.1. Multi-channel Wiener filter . . . 120
8.3.2. Spectral filter for late reverberation . . . 120
8.3.3. Coherence filter for early reverberation . . . 121
8.4. Evaluation . . . 122
8.4.1. Simulation setup . . . 122

8.4.2. Objective measures of speech intelligibility 122
8.4.3. BASSIM prediction 124
8.5. Conclusion 126

9. Conclusions **127**
9.1. Overview of achievements 127
9.1.1. Tools for the evaluation of hearing aid algorithms in realistic conditions 128
9.1.1.1. System for virtual acoustics 128
9.1.1.2. Binaural auditory system simulator 128
9.1.2. Performance of hearing aid algorithms in realistic environments 128
9.1.3. Hearing aid algorithm for improved spatial perception 129
9.2. Suggestions for improvement and future work 130
9.2.1. System for virtual acoustics 130
9.2.2. Binaural auditory system simulator 130
9.2.3. Binaural hearing aid algorithms 131
9.2.4. Perceptual evaluations 131

References **133**

Curriculum Vitae **145**

List of Publications **147**

List of Acronyms

ASW	Auditory Source Width
BASSIM	Binaural Auditory System Simulator
BTE	Behind The Ear
BRIR	Binaural Room Impulse Response
BRTF	BTE-Related Transfer Function
BSS	Blind Source Separation
CI	Cochlear Implant
CIC	Completely In the Canal
EC	Equalization-Cancellation
DRR	Direct to Reverberant Ratio
EE	Excitatory-Excitatory
EI	Excitatory-Inhibitory
HAS	Human Auditory System
HRTF	Head-Related Transfer Function
IACC	InterAural Correlation Coefficient
IC	Interaural Coherence
ILD	Interaural Level Difference
ITD	Interaural Time Difference
ITF	Interaural Transfer Function
LSO	Lateral Superior Olive
MLS	Maximum Length Sequence
MSO	Medial Superior Olive
MWF	Multichannel Wiener Filter
NC	Noise Canceler

pHAS	peripherical Human Auditory System
PSD	Power Spectral Density
RIR	Room Impulse Response
RMS	Root Mean Square
SAM	Sinusoidally Amplitude Modulated
SII	Speech Intelligibility Index
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SRT	Speech Reception Threshold
VAD	Voice Activity Detector

1. Introduction

In our modern societies more and more people are affected by different degrees of hearing losses. Recent estimates state that around 270 millions people worldwide suffer from moderate to profound hearing impairment in two ears and most live in developing countries [Tucci *et al.* 2010]. For these people, hearing aids offer at least a partial solution for their speech understanding difficulties.

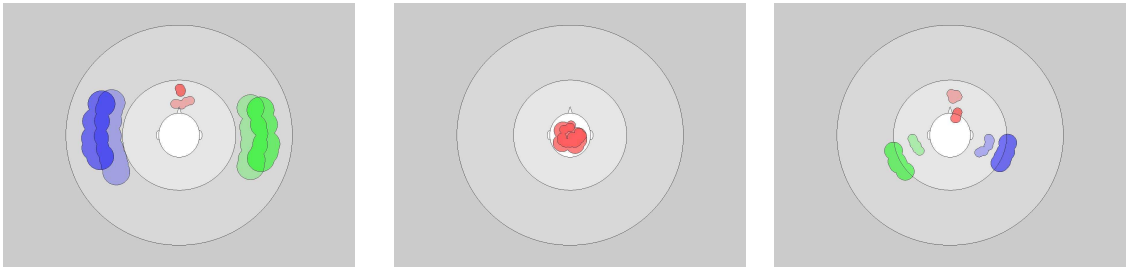
In the past decades, research on hearing aid algorithms focused on noise reduction algorithms with the aim of improving speech intelligibility and listening comfort in everyday environments. In complex listening situations however, i.e. in very noisy and multitalker environments, hearing aid users often reported that they encounter difficulties to localize correctly the sound sources of interest. The distance of sound sources, the perceived widths of auditory objects and acoustical properties of the environments are further distorted by the signal processing of the hearing devices. Moreover, it appears that with bilateral hearing aids sound sources are often perceived inside the head.

The human auditory systems uses primarily differences of the signals at both ears to build-up a spatial perceptual representation of the acoustical world. When this information is unavailable, inconsistent or distorted the spatial auditory perception can be drastically modified. In the worst case this provokes the loss of spatial qualities of the acoustical objects and a full internalization of sound. In this situation, the sensation of space is lost. The perceived sound sources have no defined positions in space anymore.

To illustrate how hearing aid algorithms modify spatial perception, a pilot study was carried out in which the spatial rendering of a selection of binaural hearing aid algorithms was investigated. In this study, the test subjects had to draw on an "answer map" how they perceived the acoustical environment. The environment was composed of a target speech signal and three surrounding incoherent noise sources placed at 90^{o}, 180^{o} and 270^{o} around the listener. The fact that the noise sources were uncorrelated created a diffuse sound field around the listeners. The room in which the experiment took place was moderately reverberant (T_{60} = 500ms). Here, the data for one listener for two different algorithms is shown in Fig. 1.1. The selected algorithms are the Blind Source Separation* (BSS) and the Multichannel Wiener Filter† (MWF) algorithms. For each condition, the test was done twice. The the responses for both test sessions are shown on the same figure.

In the Identity condition (Fig. 1.1(a)) no hearing aids were worn. This is the reference. In this condition, the listener reported hearing three independent sound sources one at each

*[Aichner *et al.* 2007]
†[Doclo & Moonen 2002]



(a) Identity

(b) Blind source separation [Aichner *et al.* 2007]

(c) Multi-channel Wiener filter [Doclo & Moonen 2002]

Figure 1.1: *Drawings of spatial perception for a speech signal located at* 0^o *and three incoherent noise sources at* 90^o*,* 180^o *and* 270^o*. Each color corresponds to a perceived different sound source.*

side, and one at the front, somewhat closer. Interestingly, the noise source played in the back was not reported by the listener. The diffuse background noise was perceived as two very diffuse sources at the sides of the listener. The target speech signal was compact and well localized and appeared closer than the surrounding noise.

Both the BSS and the MWF algorithms changed the way the listener reported his perception of the acoustical scene. For the BSS, all sounds were heard inside the head. There were no feeling of space nor defined positions for the target and noise signals. This implementation combined the signals of the left and right microphones and produced the same output for left and right speakers. This processing apparently removed the spatial cues and caused the internalization phenomenon observed here.

In this condition and for this particular test subject it appears that the MWF algorithm (Fig. 1.1(c)) offers a better spatial reproduction of the acoustical scene. As in the Identity condition (Fig. 1.1(a)), three distinct sources were perceived. The noise is also more diffuse than the target speech signal. It is however perceived more in the back and closer to the listener. For one of the two test sessions, the target signal was perceived in the head of the test subject. This indicates that even though the algorithm did a fairly good job at reproducing all the components of the scene, there were still some distortions in what the test subject perceived.

With the recent advances of wireless technology, it is possible to use collaborative bilateral hearing systems. Sharing information between the two ears allows the development of improved algorithms that exploit the binaural processing of the human auditory system. Recent research has been carried out on new binaural algorithms that include the Multi-Channel Wiener Filter [van den Bogaert 2008, Doclo & Moonen 2002], binaural beamformers, blind source separation algorithms [Aichner *et al.* 2007] or interaural coherence algorithms [Wittkop 2001, Wittkop & Hohmann 2003]. The algorithms are described in more details in Chapter 8.

There is however a lack of methods for measuring the spatial reproduction capabilities

of hearing aid algorithms. Traditional localization tests need to be more representative of everyday hearing aid experience. In this thesis, two experiments were developed with the aim of measuring objectively localization and distance perception with hearing aids in realistic conditions. They are discussed in Chapters 4 and 6. Furthermore, to increase the development and prototyping of future devices, new measures or models of spatial perception need to be introduced to reduce the burden of very time-consuming localization tests. In this thesis, the Binaural Auditory System Simulator (BASSIM) is introduced (Chapter 7). The simulator combines a model of the binaural auditory system [Breebaart *et al.* 2001] with a random forest statistical classifier [Breiman 2001] to predict how an arbitrary input signal will be perceived. BASSIM can be applied to signals processed by bilateral hearing aid algorithms and offers a prediction of source position and diffuseness based on a physiological model of the extraction of spatial cues from the input signals.

1.1. Objectives

The aims of this work are as follows:

1. Develop tools that allow the evaluation of spatial qualities of bilateral hearing instruments in realistic acoustical conditions.
2. Apply these tools to a selection of actual hearing aid algorithms.
3. Propose improvements for the future development of hearing instruments.

1.2. Thesis outline

The thesis is structured as follows:

Chapter 2: Spatial hearing
First, an overview of the human auditory system and the perceptual quantities addressed in this thesis is provided. The detection and the importance of spatial cues for the perception of source position, distance and width are discussed. A number of binaural models are introduced as well.

Chapter 3: System for virtual acoustics
In order to evaluate hearing aids in realistic conditions a system that allows the reproduction of complex acoustical environments in a plausible manner is presented. The system is based on a combination of head-related transfer functions, room simulations and head-movements reproduction. It allows the reproduction of very realistic scenes in which hearing aid algorithms can be evaluated. The publications related to this chapter are [Müller *et al.* 2011a, Grämer *et al.* 2010, Schimmel *et al.* 2009].

Chapter 4: Localization with bilateral hearing aids
The system for virtual acoustics was used to reproduce four realistic scenes: a crowded cafeteria, a busy office, a noisy street and a windy forest. In these scenes a series of

localization experiments was carried out. The system for virtual acoustics was validated and three hearing aid algorithms were evaluated. The results show that hearing aids do have a significant impact on sound localization, especially in the high rate of front-back confusions. The main findings of this chapter were published in [Müller *et al.* 2010, Müller *et al.* 2011b]

Chapter 5: Localization with bilateral Cochlear Implants: influence of head movements

Head movements play an important role in the localization of sound objects but it is not known whether bilateral cochlear implants users can use them efficiently. To test this, a localization experiment was carried out in which the test subjects had to localize speech signals of different lengths in background noise with or without head movements. The test signals varied between single words to full sentences. The system for virtual acoustics was used to track the head movements of the test subjects. The results show that bilateral cochlear implant users can exploit head movements to reduce the number of front-back confusions, provided the target signal is long enough. The limited binaural information that their hearing devices are able to transmit limit however the angular resolution of their localization performance. The outcome of the study were published in [Müller *et al.* 2011d, Müller *et al.* 2011c].

Chapter 6: Distance perception with bilateral hearing aids

Another aspect of spatial hearing investigated in this thesis was the perception of distance with bilateral hearing aid algorithms. The system for virtual acoustics was used to reproduce a large virtual reverberant room, in which sounds were simulated at various distances from the listener. Male and female speech was presented pairwise to listeners, played at a reference and a comparison distance. The task of the listeners was to detect the reference distance for different distance intervals. The test did not show any degradation of distance perception caused by hearing aid algorithms and confirms thus the results of the Speech, Spatial and Qualities of hearing (SSQ) questionnaire.

Chapter 7: Predicting spatial perception

The subjective tests described above involve the participation of a high number of test subjects and are very time consuming. To reduce this constraint, models of the binaural human auditory system can be used to predict how a given algorithm might impact spatial perception. The Binaural Auditory System Simulator (BASSIM) was developed for this purpose. It combines the binaural detection model of Breebaart et al. [Breebaart *et al.* 2001] with a statistical classifier. The BASSIM was trained on individually measured head-related transfer functions. The signals and the algorithms used for the localization experiments were processed by the simulator and the results of the subjective experiments compared with the predictions. They show good accordance, which implies that BASSIM can be used to evaluate hearing instruments.

Chapter 8: new binaural algorithm

In Chapter 8 we introduce a new algorithm that combines aspects of the binaural MWF and dereverberation algorithms. The algorithm preserves binaural cues while reducing the perceived reverberation in reverberant environments. The binaural model was used

to evaluate the performance of the algorithm and the model output was compared with the localization studies discussed above.

Chapter 9: Conclusions

The most important findings of this work are summarized in the conclusion section and suggestions for future work are discussed.

2. Binaural hearing

An acoustical scene is composed of various sound sources evolving in different positions in space. The human auditory system is able to separate effectively the sounds in different auditory objects and to characterize them with different spatial attributes: a direction, a distance and a width. As stated in the introduction, bilateral hearing aids might affect these quantities because they distort binaural cues. In this chapter, the binaural human auditory system and the perceptual quantities that are addressed in this thesis are briefly described.

Binaural hearing refers to the ability of the human auditory system to combine information from both ears by opposition to monaural hearing where only one ear is used. The binaural auditory system exploits among others differences of the signals at the two ears (the so-called binaural or interaural cues) to localize the target source precisely and segregate it from various interfering signals. The main cues available to the binaural auditory system for this purpose are the Interaural Time and Level Differences (ITDs and ILDs) at each ear [Blauert 2005, Wang & Brown 2006]. As will be shown later, some other cues are involved in localization, such as particular monaural spectral attenuations and gains due to the shape of the pinna.

The auditory space is defined by the auditory impression produced by a set of acoustical events. Fig. 2.1 gives an example of such a space, where three sources are randomly distributed in a general acoustical environment. The auditory space is divided in spatial attributes that describe a particular subjective impression in the auditory space. In this example, three spatial attributes are shown: localization, auditory source width and environment width or spaciousness. They are related to three distinctive perceptual aspects of the auditory space. The interaural cues are also related to the perception of the different attributes of the auditory space defined in Fig. 2.1 [Blauert & Lindemann 1986, Mason 2002]. This chapter is structured as follows: first a description of the peripherical human auditory system is given. Binaural localization and the role of interaural cues are briefly discussed. The concepts of auditory source width and the role of the interaural coherence are then introduced. Finally, later processing stages of interaural cues and models of the binaural system are discussed.

2.1. The peripherical Human Auditory System

The peripherical Human Auditory System (pHAS) can be divided into three different parts called outer, middle and inner ear, that correspond to the three stages of conversion of air pressure fluctuations into neural impulses. In Fig. 2.2, a schematic drawing of the pHAS is given.

In the outer ear, the auditory information is composed of sound pressure waves that arrive

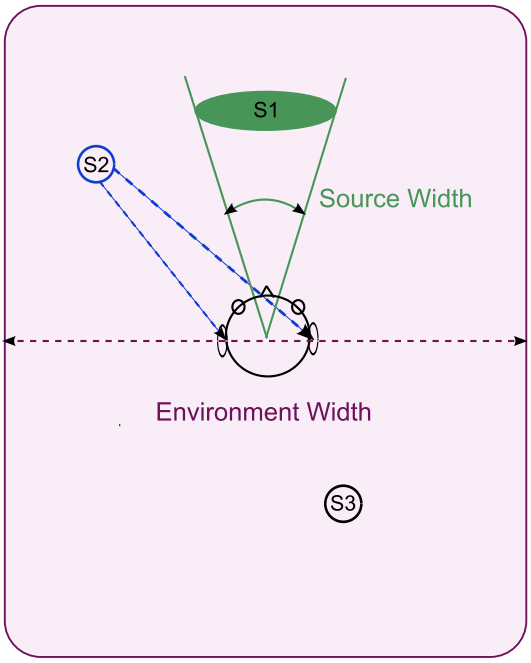


Figure 2.1: *Representation of three perceived spatial attributes in a general auditory space. S1, S2 and S3 symbolize three arbitrary sources. Localization (blue), auditory source width (green) and environment width are the spatial attributes.*

at the eardrum through the outer ear canal. The function of the eardrum is to transform the air pressure fluctuations into mechanical vibrations that are transmitted to the cochlea through a series of small bones (hammer, anvil and stirrup) at a region called oval window. The hammer, anvil and stirrup compose the middle ear.

The inner ear consists mainly of the cochlea. Its function is to transform the mechanical vibrations into neural impulses that travel to the brain through the auditory nerve. The cochlea is shaped like a snailhouse and contains a fluid. The vibrations transmitted to the cochlea induce traveling waves from the oval window along the basilar membrane, depending on the frequency content of the signal. The basilar membrane contains sensory cells (hair cells) that transform the vibrations of the fluid into neural impulses. For different frequencies, different regions of the basilar membrane are excited. When a single frequency stimulus is presented to the ear, the traveling wave induced in the cochlea propagates from the oval window along the basilar membrane and enters in resonance at a certain point in the cochlea, producing a peak response from the sensory cells at that location. A single frequency stimulus excites a region of a certain width of the basilar membrane.

From a signal processing point of view, the frequency-to-space transformation that happens in the inner ear can be seen as a filter-bank decomposition of the original signal by a high number of overlapping bandpass filters. The bandwidths of the filters are called critical bands. The widths of the critical bands determine the amount of masking and the frequency

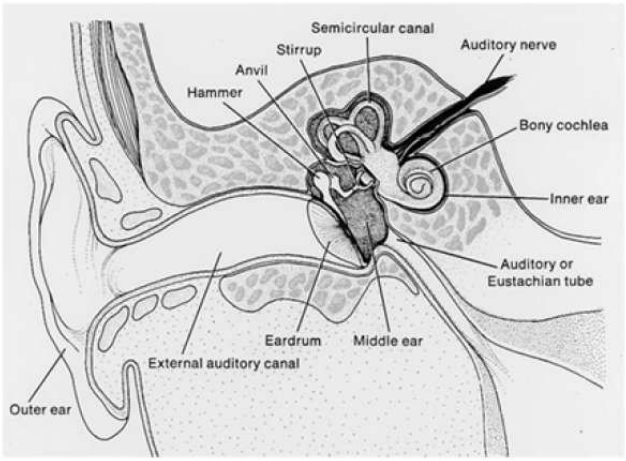


Figure 2.2: *View of the periphery of the Human Auditory System [L.E. Kinsler & Sanders 2000].*

discrimination ability of the HAS. Covering the entire range of audibility the bandwidth of the critical bands vary in size [Zwicker & H.Fastl 1990]. The bandwidths of the critical bands are assumed to be constant under 500Hz, but increase as the center frequency of a critical band increases. This implies that the HAS has a better frequency resolution at lower frequencies. Fig. 2.3 shows the shapes of the critical band filters as modeled by a gammatone filterbank. We see that as the frequency increases, the bandwidth of the filters increases too. The excitation pattern resulting from a single frequency stimulus at 2650 Hz is shown at the bottom graph. It is obtained by filtering the impulse by the different critical band filters. What is noticeable is that higher frequencies are more masked than lower ones*. This is due to the increase in bandwidth of the different critical band filters when their central frequency increases.

The range of audibility is limited from circa 20Hz till 20kHz [Zwicker & H.Fastl 1990]. The sensitivity of the HAS is not uniform over the whole frequency range. Fig. 2.4 shows the *equal loudness contours* for different sound pressure levels. They were determined experimentally by H. Fletcher and W.A. Munson and illustrate how loudness is perceived over the range of audibility at different presentation levels. ISO 226 describes the standardized and refined equal loudness contours. The perceived loudness is measured in phon. The phon is defined as the difference in sound pressure level between the equal loudness curves and the just noticeable sound at 1 kHz. The lowest curve is the threshold of hearing. By definition its corresponding loudness is 0 phon. It is the lowest sound pressure level the HAS is able to perceive. The human ear has a higher sensitivity at approximatively 4 kHz. This is due to the fact that a pressure wave with frequencies close to 4 kHz enter in resonance in the outer ear canal. The range of speech corresponds to frequencies lying between circa 130 and 8000 Hz [Zwicker & H.Fastl 1990] and coincide with the more sensitive frequency region of the HAS.

*This effect is often called the upward spread of masking.

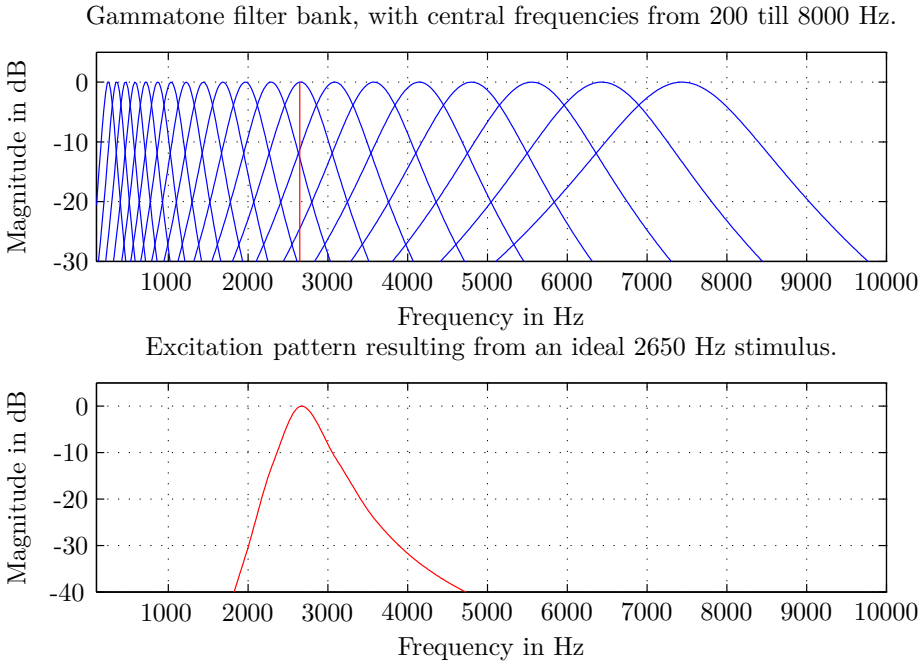


Figure 2.3: *Critical band filters based on the gammatone model. In red, an ideal 2650 Hz single frequency stimulus (on top) and the excitation pattern resulting from it (lower graph).*

2.2. Binaural Localization

The remarkable ability humans have to localize sound sources in various different conditions has intrigued scientists over many decades. A first theory on sound localization based on binaural cues was proposed by Lord Rayleigh in 1907 already. His *Duplex Theory* relates the ITDs and ILDs to the localization of sound sources. For sources in the free field, he states that the interaural time differences are dominant for localization at low frequencies whereas the interaural level differences are preponderant for localization at high frequencies. The transition from ITDs to ILDs occurs gradually at $f_c \cong 1.5$ kHz. The Duplex Theory is based on the observations that, as the maximal interaural delay for a human head of average size is 660 μs, the ITDs of periodic signals can be unambiguously decoded only at frequencies lower than f_c and that the effects of head shadowing are highest at high frequencies [Wang & Brown 2006].

Despite its simplicity, the Duplex Theory has been verified in many different discrimination experiments for a wide range of stimuli such as pure tones, clicks, broadband noise, etc. [Macpherson & Middlebrooks 2002]. The loss of phase-locking of the inner hair cells in the human cochlea at frequencies higher than approximatively 1.5 kHz does not allow the human auditory system to follow the fluctuations of the fine structure of a signal. Above this frequency, only envelope information is transmitted to later processing stages in the human brain. This property of the cochlea further supports the Duplex Theory.

The interaural time and level differences measured on a human subject are shown in

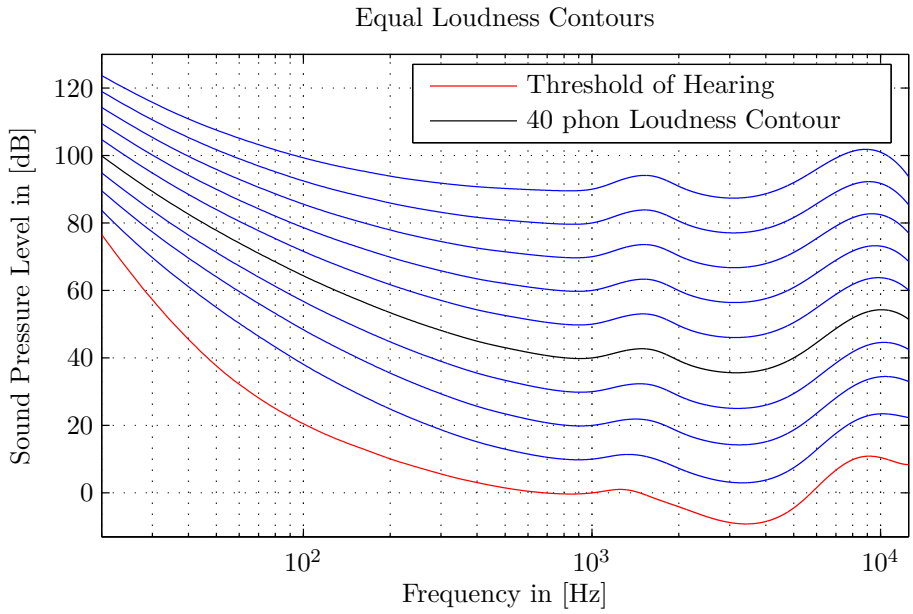


Figure 2.4: *ISO 226 equal loudness contours for different presentation levels (0 till 90 phon) [L.E. Kinsler & Sanders 2000].*

Fig. 2.5. The figure clearly illustrates the variation of interaural information with respect to the angle of arrival of sound. For this particular subject, the ITDs have a maximal value of around 620μs at $\pm 90^o$. The ITDs are symmetric around the $0^o - 180^o$ axis. The ILDs show similar characteristics. They reach however their peak of 27 dB at $\pm 105^o$. The orientation of the pinna towards $\pm 105^o$ explains this difference.

In localization studies where the test subjects are not allowed to move their heads, it was observed that normal hearing listeners have difficulties to distinguish sounds played in the front from sounds played from the back. This is generally explained by the fact that the interaural time and level differences are identical in the front and in the back for various positions in space, the cone of confusion [Wightman & Kistler 1999]. In this case the only elements available for making this distinction are pinna and visual cues. Disregarding visual cues, the spectral filtering introduced by the shape of the outer ear affects sound differently depending on whether it is played in the front or in the back. Fig. 2.6 shows curves of equal ITDs and ILDs based on HRTF measurements on human subjects. The figure shows that for a broad range of positions the interaural cues are identical.

The importance of the outer ear for discriminating sound from the front and the back is illustrated in Fig. 2.7. The figure shows the filtering induced by the head, torso and pinna measured on a human subject for both ears in anechoic conditions. The level attenuation caused by the head shadow effect is clearly visible. The levels of the spectra for positions -60^o and -120^o are much lower than on the other side. The effect is particularly strong for frequencies above 1000 Hz. For the contralateral ear, no pinna attenuation can be seen due to the strong head shadow effect.

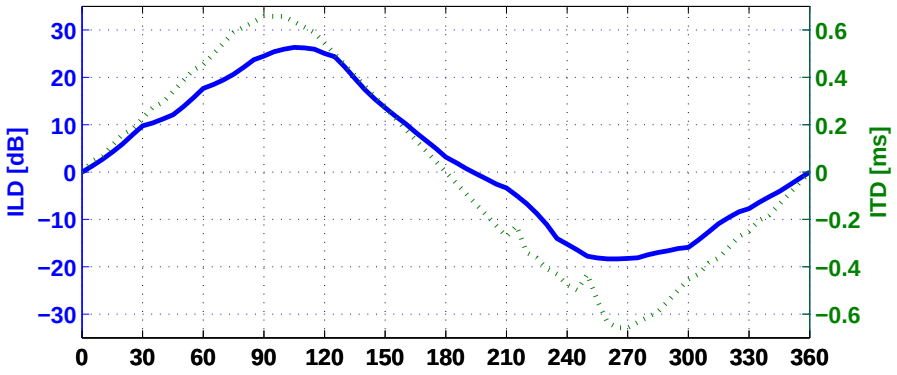


Figure 2.5: *Broadband interaural level (solid line, left axis) and time (dotted line, right axis) differences measured on a human subject in the horizontal plane.*

Providing accurate binaural information to hearing-impaired people would greatly improve their source localization and speech segregation skills. In quiet, normal hearing subjects are able to localize a source as precise as 1° in the front and around 10° on the sides [Blauert 2005]. Binaural hearing also allows a listener to selectively attend to the ear with the better signal-to-noise ratio (SNR) increasing speech intelligibility. Their ability to understand speech in realistic acoustic environments would be increased as well. A large number of studies have shown constant and significant advantage in terms of speech intelligibility for binaural hearing compared to monaural situations. This is the case for both normal-hearing and hearing-impaired subjects and even after a long absence of binaural hearing. For example, patients implanted with two independent cochlear implants (CIs) show great benefit from the implantation of the second CI, even after many years with only monaural or highly reduced binaural hearing [Ching 2005]. This motivates the development of binaural algorithms that restore at least partial binaural hearing.

2.2.1. Discrimination of Interaural Cues

Various psychoacoustic experiments aimed at investigating the salience and strength of the cues at different frequencies by presenting sounds with artificially processed interaural cues to a listener. Presented over headphones, the sound is often perceived within the head or lateralized towards one ear. In a localization experiment with sound presented via loudspeakers the sound is perceived as located outside the head.

2.2.1.1. ITD and ILD threshold for pure tones

Grantham [Grantham 1984b, Grantham 1984a] and Yost and Dye [Yost & Dye 1988] evaluated ILD discrimination thresholds, i.e. the smallest noticeable interaural intensity difference, over a wide range of frequencies for pure tones. The data they obtained is redrawn in Fig. 2.8 on the left. For clarity, the standard deviations are not shown. They usually lay between

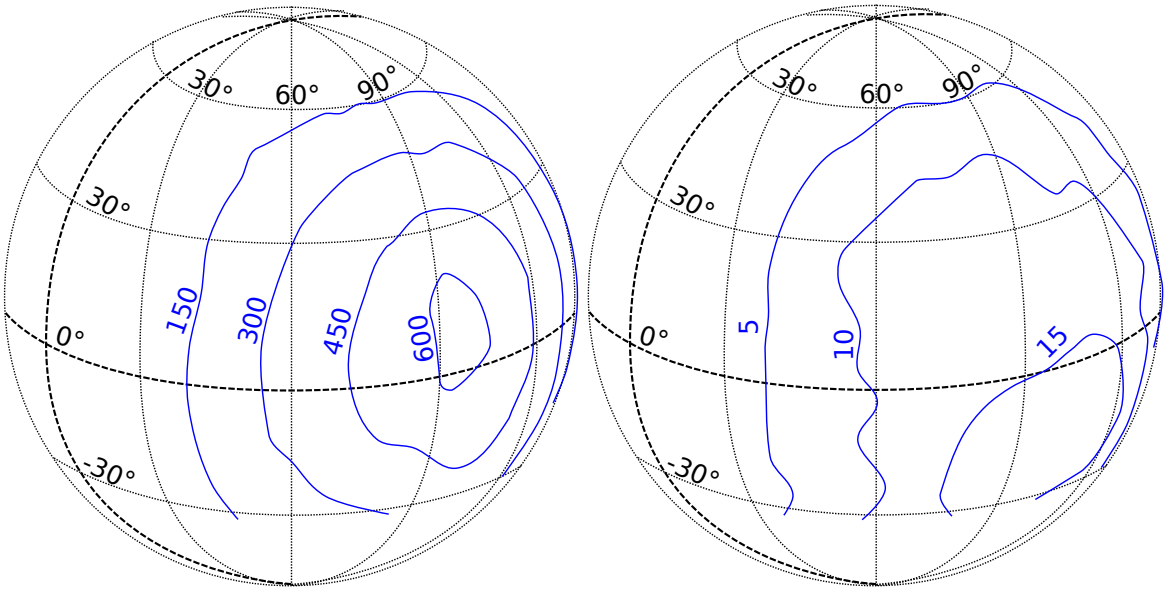


Figure 2.6: *Curve of equal ITDs (left, in µs) and ILDs (right, in dB) based on measured HRTFs. Redrawn from [Wightman & Kistler 1999].*

1-2 dB, depending on the listener and are not available for Yost's data. The data shows that the ILD thresholds are nearly constant across frequencies, except for a 1 kHz *bump*. This suggests that the human auditory system is equally sensitive to interaural level differences for pure tones covering the audible frequency range. Grantham hypothesized that the poor sensitivity at 1 kHz can be explained by different mechanism in the human binaural system for the detection of binaural cues between low and high frequency regions. As described later, there is some physiological support for his hypothesis. Yost and Dye further investigated the sensibility to ILDs for pure tones presented with different intensity differences (9 and 15 dB). As can be seen on Fig. 2.8 the thresholds they measured were worse when the signals were presented with an existing intensity difference. This can be seen as evidence for the poorer localization ability for sources presented at the side than at the front.

The human auditory system shows strong sensitivity to interaural time differences for pure tones, as can be seen in the right diagram of Fig. 2.8. ITD thresholds constantly decrease with frequencies up to 1000 Hz until 0.01 ms. In the lateralization experiment of Klump, the listeners were not able to discriminate ITDs above1500 Hz. This can be explained by the decrease of phase locking in the inner hair cells. Zwislocki and Feldman [Zwislocki & Feldman 1956] measured similar thresholds. These experimental results are consistent with the Duplex Theory.

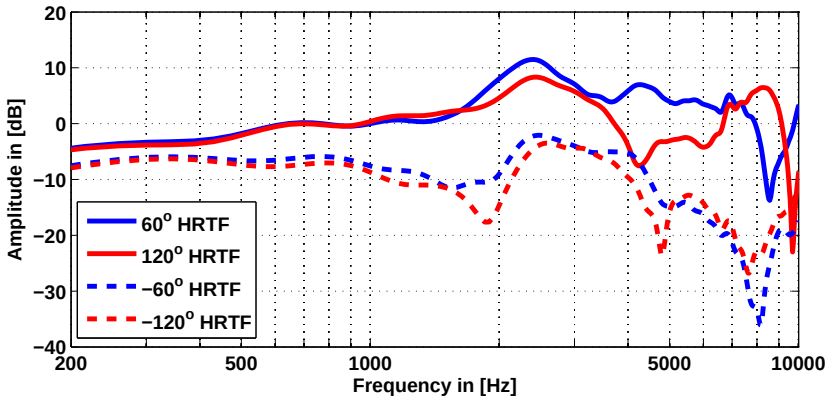


Figure 2.7: *Effect of the torso, head and pinna filtering on sound measured on a human listener for sound coming from* $\pm 60^o$ *(in blue) and* $\pm 120^o$ *in anechoic conditions*

.

2.2.1.2. Detection of high-frequency ITDs in SAM tones

However, other studies [Bernstein & Trahiotis 1994, Saberi 1998, Middlebrooks & Green 1990, Macpherson & Middlebrooks 2002] show that the human auditory system is able to use interaural time differences in the onsets and the envelopes of a signal to localize a sound source at high frequencies. With Sinusoidally Amplitude-Modulated (SAM) tones, it is possible to measure high frequency ITDs in the envelope of a signal at a single frequency, similarly as in the pure tone experiments discussed above. Bernstein and Trahiotis [Bernstein & Trahiotis 1994] reported ITD thresholds of about 0.1 ms (or 100 μs against 10 μs for low frequency pure tones ITDs) for modulation frequencies of 64 and 128 Hz (Fig. 2.9). The thresholds quickly increase with the modulation rate as the auditory system is not able to follow the fluctuations of the envelope anymore. They reported worse results as the carrier frequency of the SAM pure tones increased, with very low sensitivity for a carrier frequency of 12 kHz.

Nuetzel and Hafter [Nuetzel & Hafter 1981] examined the effect of modulation depth in ITD discrimination for SAM tones. For $f_c = 4$ kHz and a modulation rate of 300 Hz, they obtained thresholds similar to the data of Bernstein and Trahiotis (100 μs). As the modulation depth m decreases, the ITD threshold exponentially increases, reaching as far as 950 μs for $m = 0.1$. This implies that the detection of high frequency envelope ITDs in SAM tones is dependent on the carrier frequency of the tone and on the depth of modulation.

2.2.1.3. Cue discrimination experiments using other stimuli

The lateralization experiments described above all used pure tones as stimuli. Other common stimuli are clicks, narrowband and broadband noises, SAM noise, etc... For narrowband stimuli, the thresholds measured are similar to the one shown above. With broadband noise, the thresholds were slightly better for ITD discrimination [Klump & Eady 1956,

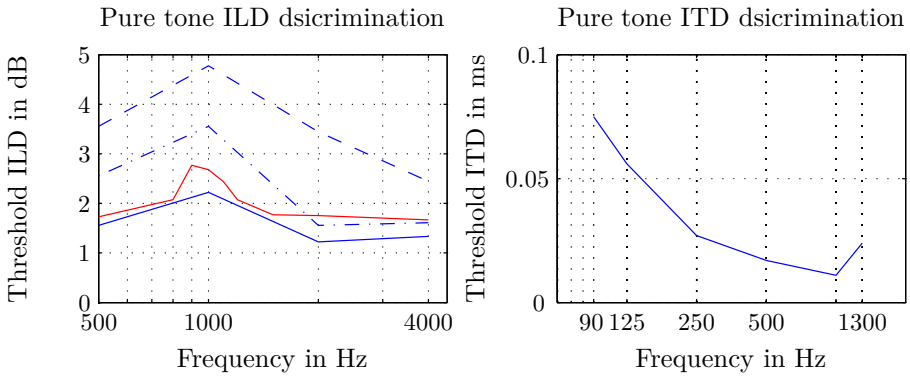


Figure 2.8: *ILD (left) and ITD (right) thresholds for pure tones over a wide range of frequencies. ILD data redrawn from [Yost & Dye 1988, Grantham 1984b]. Data from [Yost & Dye 1988] show ILD thresholds for an existing interaural level difference of 0 (blue), 9 (.-) and 15 (–) dB. ITD discrimination data is taken from [Klump & Eady 1956]. For central frequencies of 1500 Hz and above, no ITD was detected threshold was obtained.*

Bernstein & Trahiotis 1994], supporting the idea that the binaural auditory system integrates information across independent critical bands for better signal detection. The reason why the discussion presented here focused on pure tone discrimination experiments is that they allow to evaluate the relative weight of a single cue at a defined frequency in a controllable and reproducible way.

Recall that we are aiming at developing a tool that is able to objectively evaluate the loss of localization abilities due to the distortion of the interaural cues over many frequency bands. To achieve this it is essential to know the contribution of each cue in each frequency band and to understand how the human auditory system combines this information across frequencies.

2.2.1.4. Band-importance of the interaural cues for sound localization

Since it was demonstrated that the human auditory system is sensitive to both ILDs and ITDs for low and high frequency stimuli, the question may be posed whether the Duplex Theory is wrong. What is the importance of a cue in a single frequency band for the localization of a sound source? Macpherson and Middlebrooks investigated in [Macpherson & Middlebrooks 2002] the weight of the interaural cues in a *localization* experiment. The stimuli consisted of wideband (0.5-16 kHz), low-pass (0.5-2 kHz) and high-pass (4-16 kHz) gaussian noise. Using the listeners own HRTFSs, they were able to simulate through headphones sound sources from various directions. They judged the salience of each cue by adding a perceptually relevant bias on the ITDs or ILDs previously extracted from the HRTFs for each listener.

The listeners had to point to the direction from which they perceived the sound. This was done for different imposed cue biases and different signal locations. A bias weight was

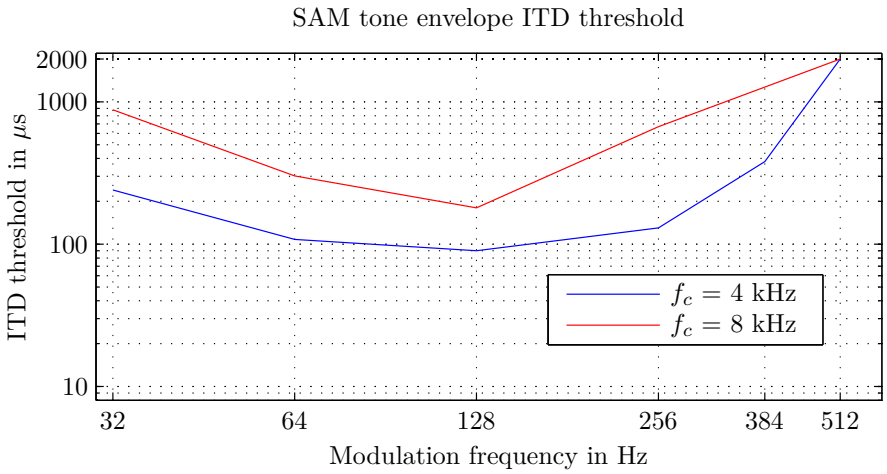


Figure 2.9: *High frequency envelope ITD threshold for fully-modulated SAM pure tones at frequencies f_c of 4 (blue) and 8 (red) kHz as function of modulation frequency. The stimuli were presented at 80 dB. Redrawn from [Bernstein & Trahiotis 1994].*

attributed based on a linear relation between the *perceived*, the *original* and the *imposed* locations. The imposed location corresponded to the theoretical position indicated by the biased cue. One cue (either ITD or ILD) was transformed at a time. A bias weight of 0 means that the perceived location corresponds to the original (i.e. the imposed bias has no influence) whereas a weight of 1 means that the perceived location corresponds to the imposed one (i.e. the source is only localized by the biased cue). The subjects in their experiments reported hearing the sound from outside their heads. According to the Duplex Theory, the imposed ITD bias should have a stronger effect for the low-pass and bandpass noise and a small effect for the high-pass noise. Fig. 2.10 shows their results. One can see strong individual differences between the bias weight measured for the different subjects.

Whereas ITDs dominate wideband and lowpass localization, they seem relatively poor for high frequency stimuli. As discussed previously, this may be due to the loss of phase-locking in the human ear above 1500 kHz. It was also shown that the human auditory system was sensitive to high frequency ITDs in the onset and the envelope of the signals.

In the same study [Macpherson & Middlebrooks 2002], Macpherson and Middlebrooks did a similar experiment with SAM highpass noise and/or stronger onsets by lengthening or shortening the duration of the onsets and offsets of the signal. They reported a slight increase of the ITD bias weight for strong onsets (mean difference 0.03). The envelope ITDs had a stronger influence on the ITD bias weight. A mean increase of 0.16 was measured. Their results showed high inter-subject differences. Listeners with an already strong sensitivity to high frequency ITDs reacted more strongly to envelope or onset ITDs than the others. Macpherson and Middlebrooks further argued that the envelope ITD cues may not play an essential role in localization based on the observation that discrimination thresholds they measured are of the order of 200-300 μs, which corresponds to a Minimum Audible Angle (MAA) of 25^o. This is insufficient to allow robust localization based on high frequency ITDs

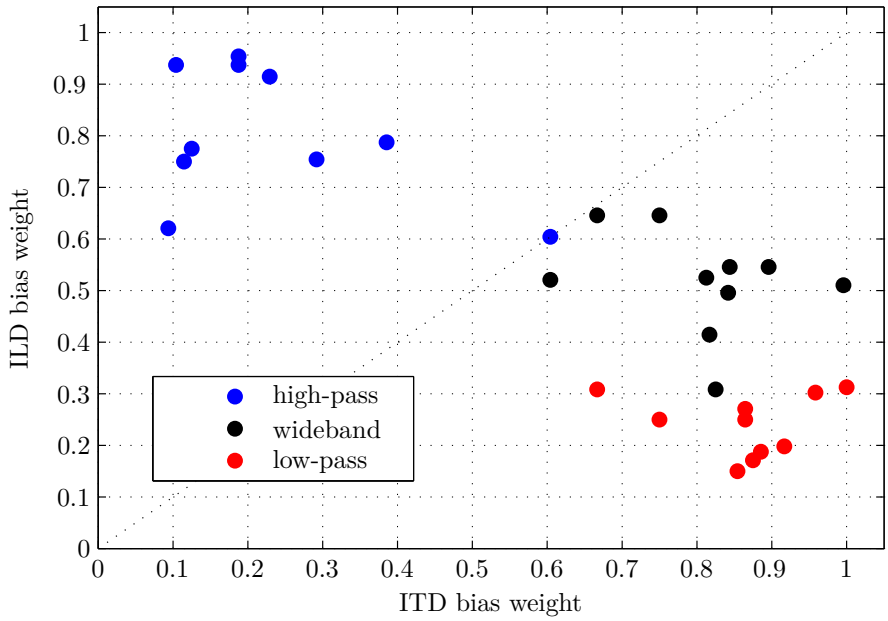


Figure 2.10: *Bias weight for 10 subjects for highpass (blue), broadband (black) and lowpass (red) gaussian noise. From [Macpherson & Middlebrooks 2002].*

alone[†]. The modulation depth actually available to the human auditory system is reduced by filtering in the peripheral auditory system. It is further decreased in a realistic environment due to diffuse reverberation. Nevertheless, envelope fluctuations that are comodulated across frequency bands play a big role in auditory grouping and source segregation, increasing speech intelligibility [Blauert 2005, Festen 1993] and have to be included in a future spatial intelligibility index.

2.2.2. Binaural cues in a reverberant environment

While the behavior and the perception of the binaural cues in an ideal anechoic situation is well known, relatively few studies have addressed how binaural cues influence spatial auditory perception in a reverberant environment. Due to interferences of single reflections, the interaural relationship of the signal at each ear fluctuates over time. Fig. 2.11 illustrates this phenomenon in an ideal simple example by considering the interference between the direct sound component and a single reflection at a wall at the right side of the listener. With a signal that consists of three tones of 480, 500 and 520 Hz the interaural cues vary over time. They are shown in the middle and right graphs in Fig. 2.11. For the interaural time differences, the strength of these fluctuations covers almost the entire range of possible val-

[†]Those are the ITD thresholds they reported in their paper for gaussian noise. For the SAM tone case discussed above, the lowest ITD threshold was 100 μs which corresponds to a MAA of 10°. A similar conclusion is valid in this case.

ues, ranging from -0.5 till 0.5 ms. It is known from room acoustics and auditory perception that spaciousness as well as the width of the perceived source are related to the amount of fluctuations [Blauert & Lindemann 1986, D.Griesinger 1992, Griesinger 1998, Mason 2002].

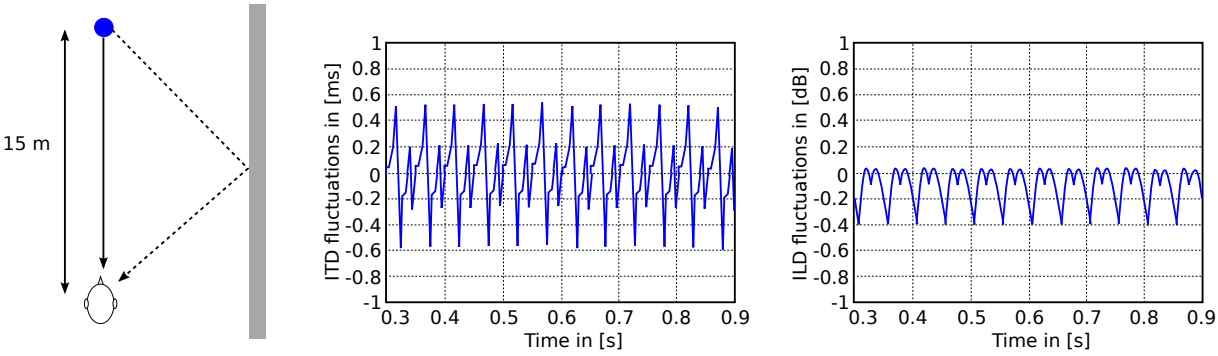


Figure 2.11: *Fluctuations in interaural time and level difference for a signal composed of three pure tones of 480, 500 and 520 Hz. From [D.Griesinger 1992, Mason 2002].*

2.2.2.1. Precedence effect

In complex listening situations with the presence of a high number of reflections, the human auditory system is still able to localize correctly a sound source. The precedence effect, or the law of the first wavefront, describes the mechanisms used by the human auditory system to do this task. The precedence effect cannot be seen as a suppression of early reflections, as they still influence the perceived width of the sound source. It is more a locking of the human auditory system on the direction pointed by the first sound components reaching the ears, the direct sound.

Precedence effect studies are usually conducted in simple settings where the stimuli consist of a lead and a lag separated by a couple of milliseconds. The lag usually comes from an other direction. A review on precedence effect studies can be found in [Litovsky *et al.* 1999]. The precedence effect can be divided into three phenomena: the fusion, localization dominance and discrimination suppression. The fusion indicates that for a lag smaller than the echo threshold, the lead and the lag are perceived as a single auditory event. The echo threshold varies between 2 and 50 ms depending on the type of stimulus.

The term localization dominance describes the fact that for delays larger than 1 ms, the perceived location of the auditory event is defined by the lead sound. For shorter interstimulus delays, the perceived position is a weighted contribution of the positions of the lead and the lag, depending on the delay between the two. It has been shown that increasing the delay shifts the perceived position to the lag.

Experiments on discrimination suppression investigate the ability of the human auditory system to detect changes in the direction of the lead and the lag. Studies [Wang & Brown 2006, Litovsky *et al.* 1999] have often assumed that the precedence effect involves an inhibitory mechanism that suppresses the activity produced by the subsequent

reflections. More recently however, Faller and Merimaa [Faller & Merimaa 2004] suggested that the interaural coherence is involved in discrimination suppression. In their model, the binaural cues are selected only in time instants and in frequency bands with high interaural coherence. Since reflections reduce the interaural coherence, the models only select interaural cues when direct sound components are present. Thus, the perceived sound source position corresponds to the direct sound component of the binaural signal.

2.3. Auditory Source Width and Spaciousness

A sound source does not only radiate from a single point in space. The sound originates from a physical object with a given shape and volume. The *Auditory Source Width* (ASW) is a concept taken from the field of room acoustics that relates the perceived volume of the source to its physical characteristics based on acoustical stimulation only. In a closed space, the impression of ASW is correlated with the set of discrete early reflections that arrive at each ear.

The presence of reflections reaching the ears from virtually all possible directions creates a diffuse soundfield. This diffuseness produces a feeling that can be described as the sensation of being "enveloped" by the soundfield or "inside the room". The term *Environment Width*, also called *spaciousness*, describes this sensation. A total loss of spaciousness results in the perception of the auditory sources inside the head. This uncomfortable feeling is to avoid for an optimal use of spatial auditory information. Environment width is mainly related to the set of diffuse late reflections.

Both ASW and spaciousness depend on the interaural relationships of the signals at each ear as was shown in 2.2.2. Therefore, information can theoretically be extracted from the binaural signals to predict the perception of the auditory space.

2.3.1. Interaural cues and measures of spaciousness

Classical theories of acoustics provide measures of the spatial attributes presented in Fig. 2.1 and relate them to auditory impressions. Most of the existing measurements are based on the room impulse response (RIR), which describes the transfer function for a source-receiver pair in a specific room. However, since we are aiming at assessing the quality of the signal presented by bilateral hearing aids, we do not have direct knowledge of the RIRs but rather have to work on the binaural signals directly.

Traditional measures of ASW and environment width are inversely proportional to the early and late InterAural Correlation Coefficient (IACC). In highly reverberant and very diffuse environment, the IACC will be low and the perceived spaciousness high. Eighty milliseconds after the arrival of the direct sound, the reflections are generally considered to affect mainly the environment width. In [Hidaka *et al.* 1995], the measures based on the IACC are defined as:

$$ASW = 1 - IACC_E = 1 - \left(\frac{\int_0^{80\text{ms}} s_L(t) s_R(t) dt}{\left[\int_0^{80\text{ms}} s_L^2(t) dt \int_0^{80\text{ms}} s_R^2(t) dt \right]^{1/2}} \right) \tag{2.1}$$

and

$$EW = 1 - IACC_L = 1 - \left(\frac{\int_{80\text{ms}}^{\infty} s_L(t) s_R(t) dt}{\left[\int_{80\text{ms}}^{\infty} s_L^2(t) dt \int_{80\text{ms}}^{\infty} s_R^2(t) dt \right]^{1/2}} \right) \tag{2.2}$$

where $s_L(t)$ and $s_R(t)$ are the left and right ear signals respectively.

A number of studies [Blauert & Lindemann 1986, Hidaka *et al.* 1995, Mason *et al.* 2005] showed that the correlation between these measures and the subjective perception of the auditory attributes is high. They are commonly used as objective measurements for the acoustical quality of enclosed space. The human auditory system has a remarkable ability to detect incoherence. Goupell and Hartmann in a set of recent papers [Goupell & Hartmann 2006, Goupell & Hartmann 2007a, Goupell & Hartmann 2007b] investigate the relation between the detection of interaural incoherence and the fluctuations of interaural cues. For narrowband stimuli, they showed that detection of incoherence was proportional to the amount of fluctuations in interaural time and level differences between the left and right signals. It is however still matter of debate whether fluctuations in interaural cues have a strong influence on incoherence detection for more general broadband stimuli.

Blauert and Lindemann [Blauert & Lindemann 1986] related the magnitude of the fluctuations of both ITDs and ILDs to perceived spaciousness. They found similar correlations between the subjective results and the IACC-based measures as with the magnitude of ITD and ILD fluctuations. The correlation was highest ($r = 0.75$) when the metric consisted of both cues combined with equal weights.

Griesinger [D.Griesinger 1992, Griesinger 1998] investigated the role of the fluctuations of the binaural cues from an acoustical point of view. He was interested in obtaining an objective measure that faithfully describes the perceived spatial attributes in different positions within a room. Arguing that different type of signals produce different spatial impressions, he developed a metric based on recorded binaural signals and not on the RIRs. His Diffuse Field Transfer function measures the ITDs of broadband signals reaching the two ears at different position in a room.

Recently, Mason [Mason 2002] developed a measure explicitly based on the fluctuations of interaural cues. His *InterAural Cross-Correlation Fluctuations function* (IACCF) combines ITDs and ILDs in a number of narrowband channels. Fig. 2.12 is an illustration of the IACCF method.

The separation into different frequency channels of Mason's model has for aim to implement a similar decomposition to the one that takes place in the human cochlea. The ITDs are obtained as the lag corresponding to the peak of the cross-correlation of the left and right ear signals. They take values between -1 and 1 ms. Following the hypothesis that loud parts of

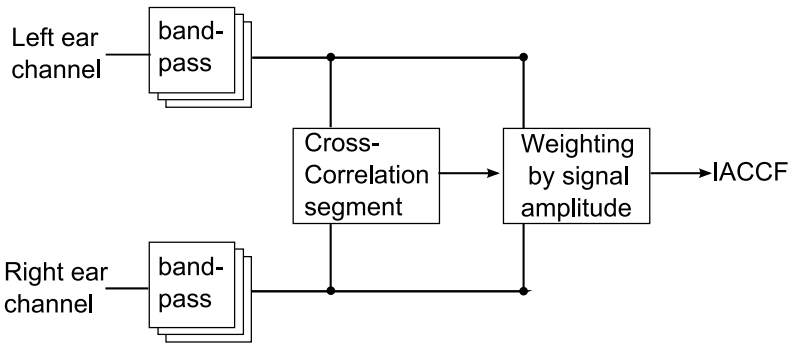


Figure 2.12: *Block-scheme of the computation of the IACCF [Mason 2002].*

the signal contribute more to spaciousness than low energy components, the cues are weighted by signal amplitude. The final value of the IACCF is taken as a mean of the amplitude of the fluctuations over each frequency channel. In [Mason 2002], Mason is aware that this might not be the optimal way of combining the cues and suggests that further research is needed to get a precise insight into how the human auditory system combines this information across frequency and time.

2.4. Sound internalization

The internalization of sound sources, or sound being perceived inside the head, is a common phenomenon that appears when sound is reproduced over headphones [Toole 1969, Sakamoto *et al.* 1976, Kim & Choi 2005, Hartmann & Wittenberg 1996]. It is also a common problem reported by bilateral hearing aid users [Gatehouse & Noble 2004]. Differences in the transmission of sound between real life and headphone listening is one of the reasons for sound internalization reported in the literature. This can create unnatural resonances and loading in the ear canal. The impossibility to reproduce small head movements also contributes to the sound being heard inside the head. The addition of artificial reverberation to anechoic recordings is also believed to increase the naturalness of the sounds, as shown by [Sakamoto *et al.* 1976].

The influence on binaural cues on sound internalization has been investigated by Hartmann and Wittenberg in [Hartmann & Wittenberg 1996]. In their study, distortions were applied to baseline interaural phase and level differences and perceptual shifts on a "inside-outside" scale were measured. The sounds were played using small speakers placed in front of the ears. The system was open, which removed the internalization effects caused by headphones discussed previously. The stimuli used in their study were artificial vowels composed of fundamental frequencies at 125 and 250 Hz and 38 harmonics. Their results show that there is a continuum between signals heard in the middle of the head and at their real position in space. The bigger the distortion to one of the interaural cues, the larger was the perceived shift on the continuum. These results suggest that the faithful reproduction of interaural information and the natural transmission of sound to the eardrum are necessary for the perfect

externalization of a sound source. This could explain why the internal sounds are so often reported by hearing aid users.

2.5. Models of the binaural auditory system

After the transformation of sound pressure waves into neural impulses by the cochlea, the auditory information is carried by the auditory nerve to various processing centers. Fig. 2.13 shows the stages of encoding of ITDs and ILDs in the superior olivary complexes (SOCs) of the human auditory brainstem. Both pathways are similar. They receive excitatory input from the cochlear nucleus (CN) and the information is primarily projected into the inferior colliculus. Physiological studies have shown that the lateral superior olive (LSO) and medial superior olive (MSO) are sensitive to changes in interaural level and time differences respectively [Yin 2002]. The MSO receives excitatory input from both the ipsilateral and contralateral sides. The LSO receives excitatory input from the ipsilateral CN and inhibitory input from the contralateral side via the medial nucleus of the trapezoidal body (MNTB in the figure).

Additionally, measures in mammals have shown that the MSO is principally composed of cells sensitive to low frequencies whereas the LSO deals mainly with high frequencies. This decomposition of binaural processing in low and high frequencies is in concordance with the duplex theory of sound [Macpherson & Middlebrooks 2002] which implies that the ITDs are extracted by the MSO and the ILDs by the LSO. The whole story is however not that simple. There is a minority of neural cells in the LSO that is sensitive to ITDs. The same is true for the ILDs.

Various models of the binaural auditory system have been proposed. Most computational binaural models can be classified by following an excitatory-excitatory (EE) or an excitatory-inhibitory (EI) model. The EE or cross-correlation models can be seen as modeling the MSO. The EI or equalization-cancelation (EC) model follows the organization of the LSO.

2.5.1. Cross-correlation models

Jeffress in 1948 already [Jeffress 1948] proposed an ITD extraction model based on a coincidence detection structure. For the same characteristic frequency (CF) a coincidence value between the left and the right ear signals is computed for different internal delays. The coincidence value is obtained after multiplication of the left and right inputs. The coincidence detector acts as a cross-correlation computation and can be seen as composed of individual EE neurons. The largest response is found at positions where the phases of the left and right signals match. The detected ITD corresponds to the lag of maximum response.

The original model proposed by Jeffress has been extended to include various frequency bands. Across-frequency processing is usually done by integrating the cross-correlation across frequency. In their weighted-image model, Stern et al. [Stern *et al.* 1988] used psychoacoustically derived weights to combine ITDs across frequency. The weights were obtained from ITD detection experiments and emphasized ITDs around 600 Hz. With their models, they could predict the outcome of various simple lateralization experiments. To deal with the periodicity

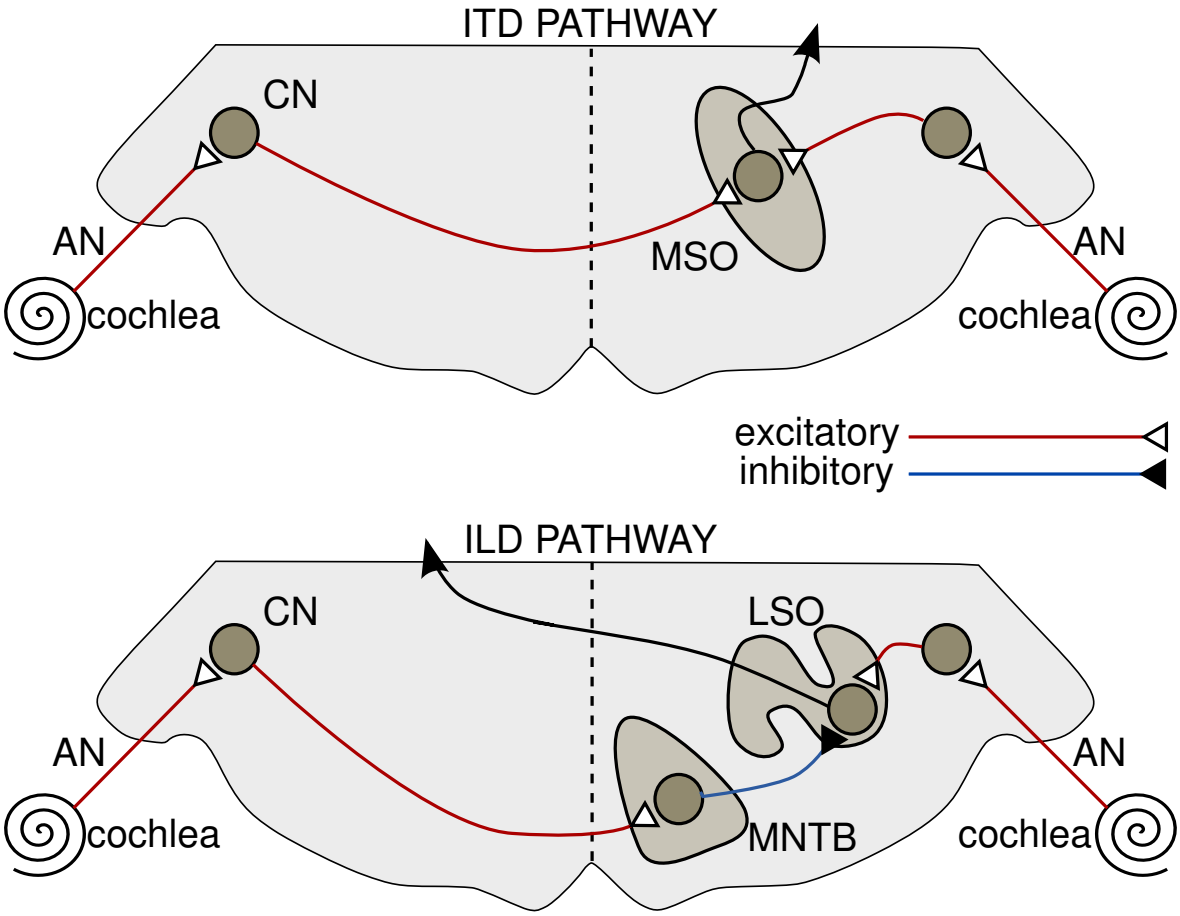


Figure 2.13: *Schematic representation of the two auditory pathway in the SOC. The processing of ITDs is shown on the top and the ILDs are shown below. Adapted from [Yin 2002].*

of the cross-correlation function, a centrality function has been proposed that emphasizes small ITDs [Stern & Colburn 1978, Trahiotis & Stern 1988]. It has been shown that the consistency of ITDs across frequencies (or "straightness") is important as well in the perceived lateralization of a stimulus [Stern & Trahiotis 2001].

Some cross-correlation models include explicitly ILD processing. To deal with ILDs, Jeffress has proposed the latency hypothesis which is based on the observation that more intense sounds tend to be initiated more rapidly than the response to low intense sounds. This effectively converts the ILDs into ITDs. In his binaural model, Lindemann [Lindemann 1986a, Lindemann 1986b]proposed an extension to Jeffress' coincidence binaural model. It includes monaural detectors and inhibition mechanisms along the τ-axis. The inhibition mechanisms are introduced to account for interaural level differences and explain aspects of the precedence effect. They are incorporated in the model as attenuator $\Delta\alpha$ along the internal delays $\Delta\tau$, following the latency hypothesis. The monaural detectors are active when the intensity of the signal is much higher at one ear than at the other (monaural listening).

2.5.2. Equalization-Cancelation models

Reed and Blum ([Reed & Blum 1990]) proposed an EI based binaural model for the extraction of ILDs. It is based on the physiological organization of the LSO. In their model, for each frequency region, the amount of excitation and inhibition from the ipsilateral and contralateral sides respectively varies along detection elements in the LSO. The ILD is detected at the element where the excitation is canceled by the inhibition. This model has been extended by Breebaart and colleagues ([Breebaart *et al.* 2001]) to include Jeffress' coincidence structure. It is composed of a discrete delay line where each tap is connected to a serie of attenuators. The structure of Breebaart's binaural processor is shown in Fig. 2.14.

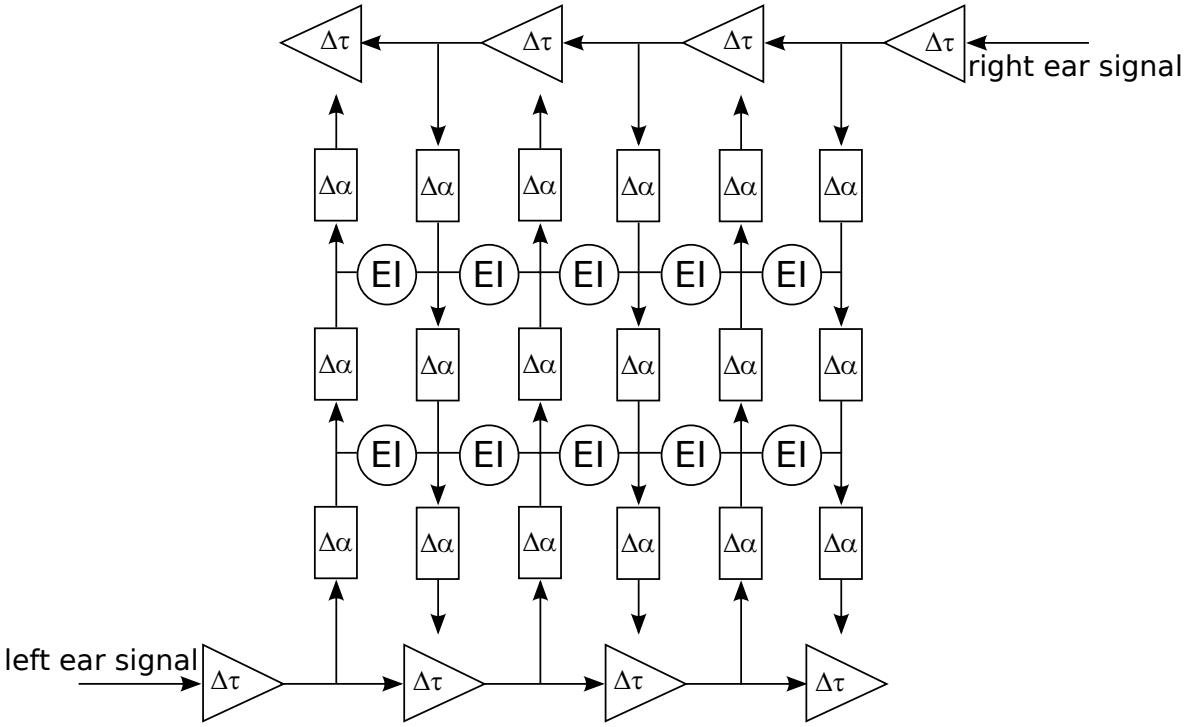


Figure 2.14: *Binaural processor of Breebaart's model composed of a delay line and a serie of attenuators [Breebaart et al. 2001].*

The outcome of the binaural processor is a two-dimensional representation of ITDs and ILDs. Breebaart's binaural model effectively computes the best ITD-ILD combination of a sound at a specific critical band as a local minimum at the output of the binaural processor. Additionally, the model takes into account the Interaural Correlation (IC) of the signal. With a low IC, the amplitude of the local minimum is higher. Breebaart's binaural model can also be seen as an extension of Durlach's EC model [Durlach 1963]. In this latter model binaural detection is done using two steps. During the equalization step, the masker components are made equal to each other to the extent possible using a delay and a gain. Detection of the target signal is achieved during the equalization stage by subtracting the signal at the two

ears after equalization. This implies that the SNR should increase at every different frequency band. This model has been used to explain binaural masking level difference experiments. It could be shown through different experiments that in most conditions, a cross-correlation based model and an EC model are equivalent (see [Breebaart *et al.* 2001]). In conditions where these predictions are not similar however, a model based on the EC mechanism is believed to give better predictions. That's the reason why an EI type of model was chosen by Breebaart. Breebaart's model has been taken as a basis for the Binaural Auditory System Simulator (BASSIM) and is described in more detail in Chapter 7.

2.5.3. Using binaural models in complex listening situations

Most of the binaural models described above were used to predict psychoacoustical data in simple acoustical environments. Faller and Merimaa [Faller & Merimaa 2004] proposed a binaural model that explicitly uses the interaural coherence for signal detection. The basic idea is that the human auditory system can rely on the interaural cues only when the interaural coherence is high. As stated previously, a high interaural coherence implies that few reflections reach the receiver and thus the direct sound component is strong. In [Faller & Merimaa 2004], the IC and ITD were computed using the cross-correlation of the left and right ear signals for a given frequency band. The model can thus be seen as a variation of the cross-correlation based binaural models. The model was able to explain aspect of the precedence effect and detect various sound sources in reverberant environment. This required however a fine tuning of the IC threshold over which the cues were defined as reliable. This threshold is dependent on the frequency band and on the acoustical environment. No automatic mechanism has been implemented to set the threshold. Moreover, the time window used for the cross-correlation computation was independent of frequency, which contradicts findings in the literature. The model could be improved by considering different mechanisms for ITD and ILD extraction by combining the EE and EI modeling approach described above. Despite its flaws, the model does well considering the complexity of realistic listening situations.

2.6. Summary

In this chapter, an overview of the human binaural auditory system was given. It was shown that the main cues used for spatial hearing are interaural time and level differences (ITDs and ILDs). They are extracted in different frequency bands and do not have the same importance across frequency. Monaural spectral cues and the interaural coherence (IC) also influence the auditory perception of space.

The auditory quantities that are discussed in this thesis were introduced and their relation to the binaural cues discussed. It was shown that in regions around the listeners where the ITDs and ILDs are ambiguous, spectral cues allow to differentiate between front and back sources. In reverberant environments, the fluctuations of interaural cues and the value of the interaural coherence define the perceived width of an auditory object. Correct reproduction of these interaural cues is necessary to remove internalization effects.

Finally, models of binaural interaction were introduced. Their aim is to explain and predict psychoacoustical data. They are mainly divided into two broad families: Excitatory-Excitatory or Cross-Correlation models (EE) and Excitatory-Inhibitory or Equalization-Cancelation models (EI). They are based on various structures of the human auditory system and are able to predict the outcome of simple binaural signal detection and localization experiments. They will be used in the later part of the thesis to predict the impact of hearing aid algorithms on spatial sound perception.

Part I.

Test methodology

3. System for virtual acoustics

3.1. Introduction

In order to investigate human auditory perception in natural conditions, sound reproduction systems that precisely simulate realistic environments were developed. The most sophisticated combine audio-visual simulations to create immersive three-dimensional environments which are undistinguishable from the real world [Lentz *et al.* 2007]. These systems typically rely on dedicated hardware, need a lot of computer power and cannot be transported to other research environments. In contemporary hearing research, there is a need for low-cost portable systems that are easy to implement and share between institutions. Despite their simplicity, they must be able to accurately reproduce the full complexity of everyday acoustical environments.

The virtual acoustics simulator discussed in this thesis combines head-related transfer functions (HRTFs) measurements with a room acoustics modeling software to generate convincing virtual environments. By means of a head-tracking sensor, the position of the head is measured and the rendering of the scene updated without noticeable delay. The system is designed to work with MATLAB on a standard Windows PC and is therefore easy to implement in other facilities.

From a perceptual point of view, head and source movements are essential contributions to a natural auditory sensation of space. The subjective experiment discussed in this chapter confirms that head movements greatly reduce the internalization phenomenon. Due to precise reproduction of head movements, most of virtual sources were perceived out of the head of the listeners. This compares favorably to traditional setups based on HRTFs.

This chapter is structured as follows: first the head-related transfer function measurement and the room simulation procedures are introduced. The real-time reproduction methods of dynamic sources and head movements are then described. The evaluation of the system is based on subjective listening tests. The test procedure and the results are discussed in the next section. Finally, brief conclusions summarize the main findings.

3.2. Virtual representation

Head-related transfer functions are used to reproduce virtual sound sources. They model the acoustic path from a source located in the free-field to the eardrum, in an anechoic environment. HRTFs contain the effects of the left and right pinnae, the head, shoulders and torso of the subject. HRTFs allow to reproduce the same sound waves at the eardrum as if they would come from the corresponding position in space. We combine HRTFs with room

simulations to generate virtual scenes.

3.2.1. HRTF measurement and calibration procedure

The sound recording and playback device is schematically depicted in Fig. 3.1. It consists of a customarily designed pair of miniature microphones and speakers located inside the ear canal. They are mounted on completely open shells of completely-in-the-canal (CIC) hearing aids and are acoustically transparent. The close location of the microphone to the speaker ($\approx$ 2mm) ensures that the sound is played at the same location where it is measured. The rigid shell forces the prototypes to sit always at the same location in the ear canal. This implies that repeated playbacks and recordings show minimal differences.

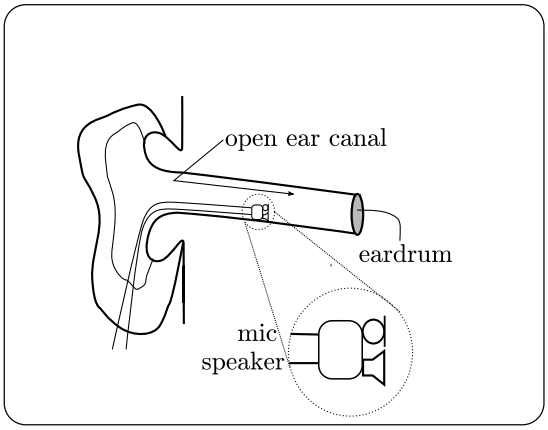


Figure 3.1: *HRTF measurement and virtual sound reproduction prototype*

According to [Kim & Choi 2005], open sound reproduction systems improve the externalization of the perceived spatial image compared to traditional closed earphones. In pilot experiments we indeed noticed that, while using the same measurement and simulation procedure, open playback guarantees a natural sensation and reduces unwanted effects such as sound internalization. The head-related transfer functions were measured using the maximum-length sequence (MLS) technique [Rife & Vanderkooy 1989].

3.2.2. Measurement and calibration of the speaker-microphone system

Let $x(t)$ be the MLS sequence played through the loudspeaker and $X(f)$ its Fourier representation. $y_l(t)$ and $y_r(t)$ are recorded by the left and right microphones, located inside the ear canal. The signals are subject to various interactions with the measurement hardware that have to be compensated. Assuming that those interactions are linear, every individual independent interfering system can be modeled by its impulse response. In the frequency domain, this results in a multiplication of the signal with the different transfer functions. For more clarity, the argument (f) and (t) as well as the side dependency will be omitted in the

following equations. The recorded signal can be written as:

$$Y = XH_sH_{\theta,\phi}H_{cic,mic} \tag{3.1}$$

where H_s, $H_{\theta,\phi}$ and $H_{cic,mic}$ are the loudspeaker, propagation and CIC microphone transfer function respectively. $H_{\theta,\phi}$ is the head-related transfer function we are interested in.

By computing the cross-correlation between the input sequence and the measured signal and by exploiting the properties of MLS signals, we obtain:

$$\hat{H}_{\theta,\phi} = H_sH_{\theta,\phi}H_{mic,cic} \tag{3.2}$$

$\hat{H}_{\theta,\phi}$ is the measured transfer function. It is subject to distortions produced by the remote loudspeaker and the CIC microphones.

A virtual sound source located at azimuth θ and elevation ϕ is simulated by convolving the left and right HRTF with the source signal x. The resulting signal is played by the CIC speaker located in the ear canal. If we use $\hat{H}_{\theta,\phi}$ as HRTF we have:

$$Y = X\hat{H}_{\theta,\phi}H_{s,cic}H_{cic,p} \tag{3.3}$$

$$= XH_sH_{\theta,\phi}H_{mic,cic}H_{s,cic}H_{cic,p} \tag{3.4}$$

where $H_{s,cic}$ is the CIC speaker transfer function and $H_{cic,p}$ models the effect of the ear canal on sound playback. $H_{cic,p}$ depends on the position of the CIC speakers and the shape of the ear canal. It is subject to strong resonances and introduces interaural and spectral distortions. The left and right CIC microphones and speakers need to be compensated for as well as they can introduce interaural phase and level differences. In this paper, we consider the loudspeaker response as part of the source and will not compensate for it.

$H_{s,cic}$, $H_{cic,p}$ and $H_{mic,cic}$ can be measured by playing the MLS sequence by the CIC speakers and recording in the ear canal:

$$Y = XH_{s,cic}H_{cic,p}H_{mic,cic} \tag{3.5}$$

$$= XH_{calib} \tag{3.6}$$

we call H_{calib} the open CIC calibration transfer function.

The *calibrated* HRTF $\tilde{H}_{\theta,\phi}$ is obtained by inverse filtering $\hat{H}_{\theta,\phi}$ with H_{calib}.

$$\tilde{H}_{\theta,\phi} = \frac{\hat{H}_{\theta,\phi}}{H_{calib}} \tag{3.7}$$

Replacing $\hat{H}_{\theta,\phi}$ in eq. 3.22 by $\tilde{H}_{\theta,\phi}$ (eq. 3.7) we have:

$$Y = X\tilde{H}_{\theta,\phi}H_{s,cic}H_{cic,p} \tag{3.8}$$
$$= X\frac{\hat{H}_{\theta,\phi}}{H_{calib}}H_{s,cic}H_{cic,p} \tag{3.9}$$
$$= X\frac{H_{\theta,\phi}H_sH_{mic,cic}}{H_{s,cic}H_{cic,p}H_{mic,cic}}H_{s,cic}H_{cic,p} \tag{3.10}$$
$$= XH_sH_{\theta,\phi} \tag{3.11}$$

This calibration method ensures that the sound played by the CIC speaker is free from distortions caused by the left and right CIC systems as shown in eq. 3.11.

3.2.3. Calibration of the BTE system

In order to evaluate hearing aid algorithms, behind-the-ear (BTE) hearing aids are simulated. Playback of the scene is done as previously through the CIC speakers. The calibration procedure is done differently. Using the same notation than in 3.2.2, we call $\hat{H}_{bte,\theta,\phi}$ the uncalibrated HRTF measured with the microphone of the hearing aid located behind the ear. Simulating the virtual sound source with $\hat{H}_{bte,\theta,\phi}$ gives:

$$Y = X\hat{H}_{bte,\theta,\phi}H_{s,cic}H_{cic,p} \tag{3.12}$$
$$= XH_{bte,\theta,\phi}H_{s,cic}H_{cic,p}H_{mic,bte} \tag{3.13}$$

where $H_{mic,bte}$ is the transfer function of the BTE microphone. In eq. 3.13, the effects of the CIC speaker and the ear canal can be seen. Applying the open CIC calibration filter (eq. 3.6):

$$Y = X\frac{H_{bte,\theta,\phi}H_{s,cic}H_{cic,p}H_{mic,bte}}{H_{calib}} \tag{3.14}$$
$$= X\frac{H_{bte,\theta,\phi}H_{mic,bte}}{H_{mic,cic}} \tag{3.15}$$

The HRTFs recorded at BTE position differ significantly from the CIC HRTFs. The strong ear canal resonances naturally present in the CIC recordings are missing in the BTE recordings which results in a large coloration difference between BTE and remote loudspeaker playback. To reduce this effect we use the diffuse field equalization method as described in [Moeller 1992]. The diffuse gain is obtained by averaging the HRTFs over all measured positions. We define the diffuse calibration filter as:

$$H_{calib,bte} = \frac{H_{mic,cic}\frac{\sum_{\theta,\phi} H_{cic,\theta,\phi}}{N}}{H_{mic,bte}\frac{\sum_{\theta,\phi} H_{bte,\theta,\phi}}{N}} \tag{3.16}$$

Applying the diffuse calibration filter to eq. 3.15 we get:

$$Y = X\frac{H_{bte,\theta,\phi}H_{mic,bte}}{H_{mic,cic}}H_{calib,bte} \tag{3.17}$$

$$= XH_{bte,\theta,\phi}\frac{\frac{\sum_{\theta,\phi} H_{cic,\theta,\phi}}{N}}{\frac{\sum_{\theta,\phi} H_{bte,\theta,\phi}}{N}} \tag{3.18}$$

In eq. 3.18 the effects of the measurement microphones are removed. The use of the diffuse gains cancels the coloration differences introduced by different playback and measurement positions.

3.2.4. HRTF interpolation

HRTFs are typically measured for a limited number of sound source positions on a sphere around the subject. The resulting spatial sampling is usually dense enough so that convolving direct sound and surface reflections from arbitrary directions with their nearest-neighbour HRTF will maintain perceptual accuracy. However, it may be too coarse to render small head movements and smooth sound source displacements without producing audible artifacts. By interpolating HRTFs between measured positions, the spatial resolution can be artificially increased until sufficient accuracy is achieved.

In general, interpolation could be done in the time or in the frequency domain and one could use one of several standard interpolation methods like linear, sinc or spline interpolation. Those methods however show a poor performance in a mean-square error (MSE) sense as well as in subjective listening tests. It has been shown that the performance of interpolation in the time or frequency domain can be improved by compensating HRTFs prior to interpolation according to the time of arrival of sound [Christensen *et al.* 1999, Matsumoto *et al.* 2004a]. That is, the HRTFs are time aligned and interpolation is carried out on the time-aligned HRTFs. In order to achieve sub-sample precision in the time alignment, the time of arrival itself is also interpolated. For the interpolation of the time-aligned HRTFs, standard interpolation techniques like linear, spline and sinc interpolation were compared and the best results are obtained using linear interpolation [Matsumoto *et al.* 2004a]. In this work, HRTFs are measured for 12 equal spaced positions in the horizontal plane which is equivalent to an angular resolution of 30^o. We interpolated the set of HRTFs between adjacent positions to an angular separation of 1^o. We used time-aligned linear interpolation to obtain the missing HRTFs.

3.2.4.1. Room modeling and simulation

Reproducing virtual sources with HRTFs alone results in a dry and unnatural percept. Human beings are rarely in spaces free from reflections. It has been shown that reflections help for the externalization of sound [Kim & Choi 2005]. With a perfect room simulation software it is theoretically possible to reproduce any reflective space and conduct realistic perceptual experiments.

The room simulations were done using the freely available ROOMSIM software [Schimmel *et al.* 2009]. ROOMSIM is an advanced MATLAB toolbox that combines both specular and diffuse reflections for the generation of perceptually accurate Binaural Room Impulse Responses (BRIRs). ROOMSIM uses an efficient implementation of the diffuse rain algorithm REF associated with the classical image source model to reproduce the stochastic and deterministic characteristic of room impulse responses in shoebox-type rooms. The direct sound and every successive reflection are associated to the closest transfer function of the HRTF catalogue.

Before playback, the resulting impulse responses need to be calibrated in order to compensate for the ear canal resonances, the microphones and the speakers of the open CIC system. The calibration is done by inverse filtering the BRIRs with the respective impulse responses. In a frequency domain notation, the final transfer function from the source to the receiver for one ear, $H_{s,r}(f)$ is given by

$$H_{s,r}(f) = \frac{R_{s,r}(f)H_{HRTF}(f)H_m(f)}{H_{calib}(f)} \tag{3.19}$$

where $R_{s,r}$ is the room transfer function from the source to the receiver, H_{HRTF} the corresponding HRTFs, H_m the open CIC microphone used for the recording of the HRTFs and $H_{calib}(f)$ the calibration transfer function. H_{calib} is obtained by measuring the transfer function from the open CIC speaker to the microphones and characterizes the open CIC system positioned inside the ear canal. It is composed of the transfer functions of the speaker (H_s), the microphone (H_m) and of the ear canal (H_c). H_{calib} is measured by means of the same MLS correlation procedure that is used to measure the HRTFs.

The signal sent to the left and right speakers is obtained by convolving the input signal ($x(t)$) with the respective impulse response. Considering the effects of the open CIC system and the ear canal, the sound wave at the eardrum of the listener ($y(t)$) can be written as:

$$Y = XH_sH_cH_{s,r} \tag{3.20}$$

$$= X\frac{H_sH_cR_{s,r}H_{HRTF}H_m}{H_sH_cH_m} \tag{3.21}$$

$$= XR_{s,r}H_{HRTF} \tag{3.22}$$

where the frequency variable (f) has been omitted for clarity. Eq. 3.22 indicates that the calibration procedure effectively removes any differential differences due to the left and right

open CIC systems and to the ear canal resonances.

3.3. Head-tracking and dynamic scene rendering

The real world is dynamic. Sound sources move along determined trajectories in space. Additionally, listeners constantly use small head movements and take advantage of the changes in acoustical spatial cues for building an auditory representation of the situation they are in. The interactions between the listener, the source and the environment need to be reproduced with minimal latency to be perceptually plausible. In such dynamic scenes, the BRIRs are continuously changing according to the source and the receiver positions and orientations. In [Savioja *et al.* 1999], Savioja et al. classify existing systems for virtual acoustics following two implementation strategies: "direct room impulse rendering" and "parametric room impulse rendering".

Systems that follow the direct room impulse rendering implementation use pre-generated BRIRs covering a grid of defined listening positions for the listener and the sound sources. For the generation of sound, the system filters the input signal with the BRIR corresponding to the position of the source. Depending on the complexity of the scene and the source and listener trajectories, this technique needs a lot of storage space and computer memory.

In the parametric room impulse rendering method on the other hand no BRIRs are pre-computed. The impulse responses are generated in real time following room acoustic models. These models commonly separate the room impulse response into direct sound, early reflections and late reflections components. The rationale behind it is that the direct sound and the early reflections show a deterministic and time-variant behavior depending on the positions of the source and the receivers and the geometry of the room. Late reflections however are considered as stochastic and slowly varying in time. Furthermore, according to classical room acoustics theory [Griesinger 1998], they determine different perceptual quantities. Direct sound and early reflections define characteristics of the sound source while late reflections contribute to the sense of "space" and characterize the environment of the scene. Practically, this implies that for real-time sound generation, the early components of the BRIRs need to be constantly updated whereas the late reflections can be predetermined. Efficient existing implementations work on dedicated DSP-based platforms and require a real-time operating system with a scheduler that can guarantee processing deadlines (see for example [Gardner 1995]). A standard PC cannot satisfy these requirements,

In order to fulfill the portability requirements and remain perceptually accurate, our system combines aspects of both techniques. Similarly to direct room impulse rendering techniques, BRIRs covering all possible source and listener positions for a given scene are pre-generated and stored in memory. Additionally, the modeling principles of the second category are exploited to satisfy the real-time requirements caused by head-movements. The algorithm is described in the following sections.

3.3.1. Generating moving sources

For the rendering of dynamic scenes without support for head movements, there are no real-time requirements and the signal processing is straightforward. For a sound source which is moving at a constant speed on a circle around the listener, one could either define a spatial resolution or a temporal resolution determined by the speed of the source. We set a fixed temporal resolution and restricted also the spatial resolution, that is, we used fixed nearest-neighbor positions of the actual sound source to generate the impulse response. Additionally, we tried linear interpolation of the two nearest neighbor positions. The processing of the audio signal was done in a block-wise manner as illustrated in Fig. 3.2.

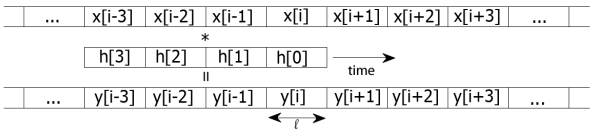


Figure 3.2: *Block-wise signal processing. $x[i]$ denotes an input signal block i, $y[i]$ an output signal block i and h is the impulse response divided into N blocks. In this example, $N = 4$. All blocks are of the same length l.*

The processing of moving sources implies a source position dependent impulse response h_θ. Since the trajectory of the moving source is known in advance, the position dependency can also be formulated as a time dependency. The output signal in the case of a moving source is obtained by:

$$h[k] = \sum_{n=0}^{N-1} \sum_{m=nl}^{(n+1)(l-1)} x[k-m]h_\theta[m] \tag{3.23}$$

where the indices k and m address single samples of the signal and the impulse response respectively, x denotes the input signal, N the number of blocks the impulse response is divided into and l is the block length. h_θ is the position dependent impulse response, calculated using Eq. 3.22. Every input block is filtered with the impulse response corresponding to the source position at the time when this very block is played. The inner sum is nothing else than the convolution of an input signal block with a block from the impulse response. Reformulating Eq. 3.23 in terms of block processing yields:

$$y[i] = \sum_{n=0}^{N-1} x[i-n] * h_\theta[n] \tag{3.24}$$

where $x[i]$ is the input signal block i and $h_\theta[n]$ a block from the impulse response. In the following, we will always use the block processing notation. This implies that the time is discretized into intervals with the same duration as a block, especially, the variable t denotes not the continuous time but the "time index" in the "unit" [block]. A more comprehensive graphical representation of the processing is depicted in Fig. 3.3.

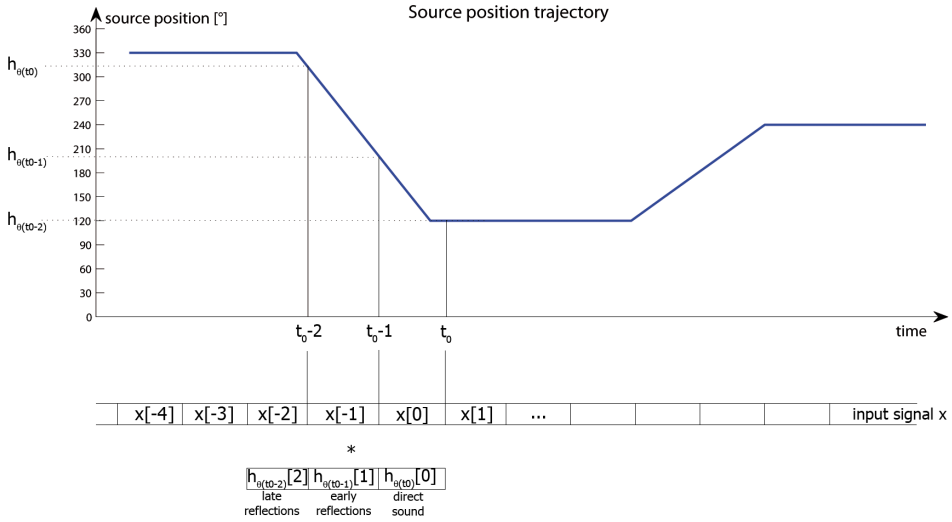


Figure 3.3: *Block-wise signal processing for dynamic scenes. The impulse response $h_{\theta(t)}$ is determined by the source position θ at time t. The input signal x is processed in blocks $x[-2], x[-1], x[0], \ldots$. The impulse response h is also divided into blocks of equal length ($h[0]$, $h[1]$, $h[2]$, ...), representing different properties of the room like direct sound, early reflections, late reflections and reverberant tail. For simplicity, h has only a length of three blocks in this example. In the real system, the impulse response is much longer. The output signal block $y_{t_0}[0]$ at time t_0 is given by $y_{t_0}[0] = h_{\theta(t_0)}[0] * x[0] + h_{\theta(t_0-1)}[1] * x[-1] + h_{\theta(t_0-2)}[2] * x[-2]$. Note that every input signal block is filtered with the impulse response corresponding to the source position at the time when this very block is played.*

To increase the efficiency of the processing, the impulse response is truncated to an integer multiple of the block size l so that N is also an integer. This means that we are losing a part of the reverberant tail of the impulse response. For a typical configuration with a block size of 512 samples and a sampling frequency of 44.1 kHz, we truncate the impulse response by at most 511 samples. If the position of the sound source is changing, h_θ changes instantaneously. No fading mechanism was implemented.

3.3.2. Updating head movements

As described in section 3.3.1, the rendering of dynamic scenes with pre-computed BRIRs can be done completely offline. The extension of the system to compensate for head movements is a more challenging task since we have no prior knowledge about the head movements of a subject. We can still use pre-rendered impulse responses, but the audio signal has to be filtered in real-time. From experimental listening tests with moving sources, we know that we need a spatial resolution of at least 1° for the offline calculated impulse-responses. The block-wise processing from the dynamic scenes was kept, the block size defines the size of the audio signal buffer and therefore also the processing delay and the temporal resolution. The

processing delay should be as small as possible which implies a small block size and also a high temporal resolution. For every block that is played, the impulse response is updated according to the head position that is obtained by polling the head-tracker.

The differences in the signal processing compared to the simulation of dynamic scenes are limited. Eq. 3.24 still holds, with the following differences: the impulse responses h_θ are pre-calculated for all possible receiver orientations (head positions) instead of all possible source positions. The most important difference is that the position dependent impulse response $h_\theta(t)$ should now correspond to the actual receiver orientation and not to the source position at the time when the sound was emitted. Figure 3.4 shows a graphical representation of the processing.

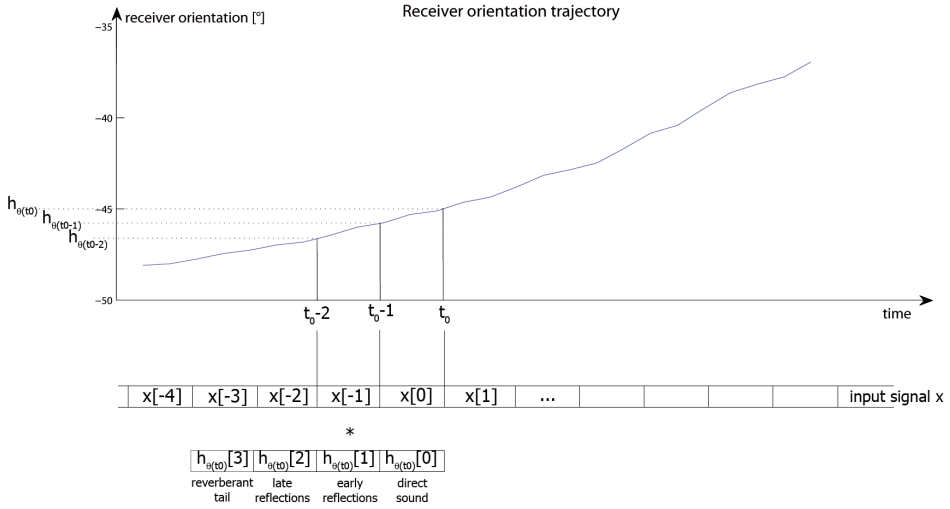


Figure 3.4: *Block-wise signal processing for the compensation of head movements. The impulse response $h_{\theta(t)}$ is determined by the actual receiver orientation at time t. Ideally, the output signal block $y_{t_0}[0]$ at time t_0 is given by $y_{t_0}[0] = h_{\theta(t_0)}[0] * x[0] + h_{\theta(t_0)}[1] * x[-1] + h_{\theta(t_0)}[2] * x[-2] + h_{\theta(t_0)}[3] * x[-3]$. The difference to the processing of moving sources is that only the impulse response $h_{\theta(t_0)}$ is used instead of composing the impulse response of the blocks $h_{\theta(t_0)}[0]$, $h_{\theta(t_0-1)}[1]$, $h_{\theta(t_0-2)}[2]$ and $h_{\theta(t_0-3)}[3]$. The real processing differs from the ideal processing which is computationally too demanding. For $a = 2$, $y_{t_0}[0]$ at time t_0 is given by $y_{t_0}[0] = h_{\theta(t_0)}[0] * x[0] + h_{\theta(t_0)}[1] * x[-1] + h_{\theta(t_0-1)}[2] * x[-2] + h_{\theta(t_0-2)}[3] * x[-3]$. The consequence of this simplification is that the late reflections of the room are simulated using a somewhat outdated impulse response.*

To simplify the calculations, we filtered the input signal only with the first part of the actual impulse response which corresponds to the direct sound and the early reflections. The late reflections and the reverberant tail are filtered with a somewhat outdated impulse response. In other words, the impulse response is divided into two parts, where the first part is used for an exact processing and the second part is an approximation tho the actual impulse

response. Let us assume that the first a blocks of the impulse response are used for the exact processing. The output signal is then given by

$$y[i] = \sum_{n=0}^{a-1} x[i-n] * h_{\theta(t)}[n] + \sum_{n=a}^{N-1} x[i-n] * h_{\theta(t-a+1)}[n]. \tag{3.25}$$

The first sum represents the part of the impulse response which is used for an exact simulation, the second sum represents the approximated parts. This means that the late reflections and the reverberant tail are filtered using an impulse response which corresponds to the position of the receiver (head) at the time $t - a + 1$. This introduces some sluggishness in the virtual acoustics system. Listening tests will show if it is audible. Eq. 3.25 does not show the simplifications that allow a faster processing. The crucial point is that in the actual implementation only the first sum has to be evaluated for every block. The second sum is a by-product of the first sum and has to be evaluated only once and not for every block.

When using this modification, a real-time processing is possible. With a block size of 512 samples (as suggested by Lentz et al. [Lentz *et al.* 2007]) and a sampling rate of 44.1 kHz, the head position is updated at a rate of 86 Hz. The overall processing delay is then 512 samples plus another 512 samples from the soundcard buffer, in total 1024 samples or 23.2 milliseconds. The impulse response was divided into two parts as described, the boundary was set to 81.3 milliseconds (81.3 ms for precise processing, 162.5 ms for sluggish processing of the reverberant tail). In pilot listening tests, no sluggishness or artifacts caused by fast head movements were noticed by the test subjects.

3.4. Perceptual evaluation

The realism of sounds generated by HRTFs has been addressed by numerous studies [Bronkhorst 1995, Wightman & Kistler 1989, Rychtarikova *et al.* 2009a]. Common limitations are poor externalization of the frontal sound image and higher number of front-back confusions, especially when non-individualized HRTFs are used for sound generation. The causes of these effects are believed to be inaccurate HRTF measurements and virtual sound reproduction and the inability of the existing systems to reproduce small head movements [Wightman & Kistler 1999].

To evaluate the perceptual accuracy of the system we decided to confront the simulation to the real world. We simulated with the ROOMSIM software a room in our laboratory. We presented sounds through loudspeakers located in the room. From the same positions in the simulated room, we presented fully virtual signals through the open CICs. The task of test subjects was to identify which of the presented sounds were virtual and to rate the degree of externalization of the sound sources. To test if the motion tracking sensor and the signal processing is fast enough to render a convincing scene where listeners cannot hear any sluggishness when listeners move their head, they had to assess the stability of the sound source as well.

The simulated room is an acoustically treated shoebox-type room with octave-band re-

verberation times (T_{60}) shown in Table 3.1. It is 6.53 meters large, 5.72 wide and 2.34 high. The receiver is set at position (3.69, 2.85, 1.15) facing the long wall. The loudspeaker ring was centered on the receiver position at a distance of 1.5 meters with an angular spacing of 30°. Real loudspeakers are not omnidirectional sources. They emit more to the front than to back. To take this is an account, we modeled the directivity of the twelve sound sources as a three-dimensional cardioid, emitting towards the receiver. This implies that most of the reflective energy comes from the floor, the ceiling and the facing walls.

f [Hz]	125	250	500	1000	2000	4000	8000
T_{60} [ms]	230	270	270	210	230	300	300

Table 3.1: *Octave band reverberation times in [ms]*

The walls of the room are composed of bended absorbent wood panels, the floor was covered by a carpet and the ceiling composed of sound treated wood panels. The frequency dependant absorption and diffusity coefficients of the enclosure were set to match the measured binaural room impulse response. The first and strongest reflections reaching the receiver are produced by the roof and the ceiling of the rooms Those reflections convey similar azimuthal information than the sound source and facilitate its localization.

A total of nine individuals served as volunteers for this listening test. There were six normal hearing male subjects, two normal hearing female subjects and one hearing impaired male subject (one of the authors). The hearing was verified by standard clinical audiometry. All of them except the hearing impaired subject had participated in earlier localization experiments and were experienced listeners. The were aged 25-48 years with a mean of 35 years.

3.4.1. Stimuli

The test stimuli consisted of speech and white noise signals. The speech material was taken from the Timit database [Garofolo *et al.* 1993]. It consists of a sentence spoken by different native male american speakers. The sounds were presented at a level of 60 dB. The intensity of the signals was roved between successive presentations by 2 dB. To avoid listeners to associate a spectral coloration to a loudspeaker, the spectrum of the noise signals was randomly colored. All sounds were bandpass-filtered between 400 Hz and 8000 Hz to remove the frequencies which cannot be reproduced by the open CIC speakers.

3.4.2. Procedure

The test was divided into 2 rounds of 16 trials plus a single training round in the beginning. The purpose of the training was to familiarize the subject with the test procedure. In the first round, 16 trials with speech signals only were presented randomly over loudspeaker or with the simulation. The signals were played from four different positions ($0°, 90°, -90°, 180°$). Every position was played twice. Eight speakers, taken randomly from the Timit corpus, were selected and presented once over the loudspeakers and once over the open CIC speakers.

In the second round, noise signals were used. The sound was repeated until the subject interrupted the sound output. The test subjects had all the time to listen to the source. The listeners were encouraged to move theirs heads but the head-movements were not compulsory. The task of the subjects was to answer the following questions:

1. On a scale from 1-5, do you hear the sound source in your head or from the loudspeaker?
2. On a scale from 1-5, does the sound source remain stable if you turn your head?
3. Where does the sound come from: Loudspeaker or simulation?
4. (If the answer to question 3 was "simulation"): Why? Any further remarks?

Questions 1 and 2 were illustrated with the Fig. 3.5. The subjects repeated the experiments at a second session. The data for the test and retest sessions show minimal differences.

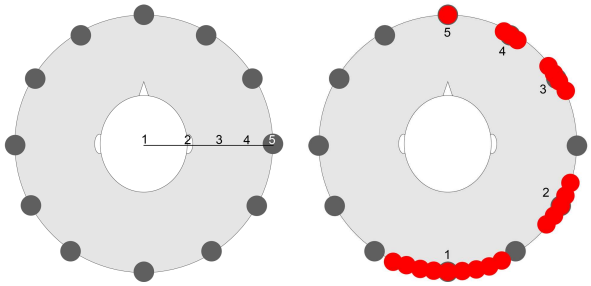


Figure 3.5: *Answer maps for questions about externalization and stability*

3.4.3. Results

3.4.3.1. Externalization

The ratings of the externality are shown in Fig. 3.6. It presents the results of all subjects, test and retest combined. The corresponding numerical values are listed in Table 3.2. Significance was tested with the Wilcoxon matched pairs signed rank test. Significance was set at $p < 0.05$. The externality rating was very high for both the speech and the noise signals. It is above 4.59 for the simulation, which is remarkable.

		Extern.	Stability	Avg
Speech	real	4.95	5	4.975
	virtual	4.59	4.88	4.735
Noise	real	4.83	4.98	4.905
	virtual	4.65	4.81	4.73

Table 3.2: *Average rating of externalization and stability*

During the tests, it appeared that sound sources presented in the front (0^o) were often perceived inside the head but as soon as the subjects moved their head, the source jumped

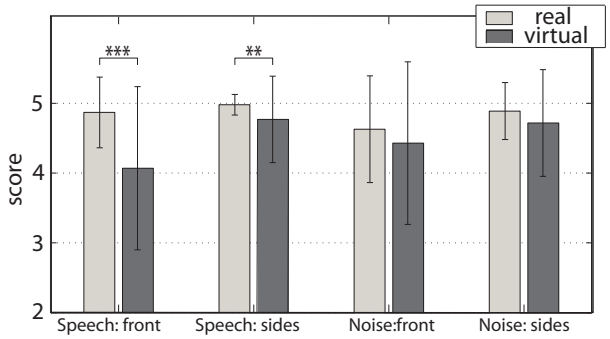


Figure 3.6: *Rating of externalization for all subjects*

out of the head and was well externalized. It remained external even if the subject returned to the starting position and kept their head still. Test subjects asked if they should rate the externalization of their first impression or the externalization after the head movement. They were told to rate the externalization during the head-moving phase. This instruction probably improved the mean externalization rating. In Fig. 3.6, the results for signals presented at the front and at sides are analyzed separately. Similarly to results reported in the literature, the externality for the frontal position was worse. For the speech signals, the difference between virtual and real playback where highly significant ($p > 0.001$) for sources at the front. Interestingly, the externality of the noise signals was similar for the two sound reproduction methods. The test subjects even reported hearing the noise inside the head for signals played by the loudspeaker in front of them. For source positions from the sides, the externality was higher. The difference between real and virtual playback was still significant for the speech signals, although at a lower p-value ($p < 0.01$). The difficulty of externalizing frontal sources can be due to the fact the human auditory system is sensitive to very small differences in interaural cues for frontal positions. Just-noticeable-differences (JND) detection experiments indeed show a higher sensitivity of the human auditory system for signals that produce a central image compared to lateralized presentations. Nevertheless, the experiment shows that head movements play an essential role for the externalization of virtually generated sound sources.

However, a source which appeared inside the head, even if only in the first few seconds of a sound presentation, was often used as a cue to detect the simulation. This can also be seen in the results of question 4.

3.4.3.2. Stability

The definition of stability is illustrated in the right part of Fig. 3.5. If the processing delays induced by the signal processing or the head-tracker would be too high, the listeners would perceive a lag and more diffusity in the simulated sources. It is known that an inaccurate reproduction of dynamic interaural cues impacts the perceived compactness of virtual sound sources.

The ratings of the stability are shown in Fig. 3.7. The corresponding numerical values are listed in Table 3.2. Even though the ratings of stability are high they are significantly different for the two conditions (4.85 for virtual sources against 4.94 for real signals).

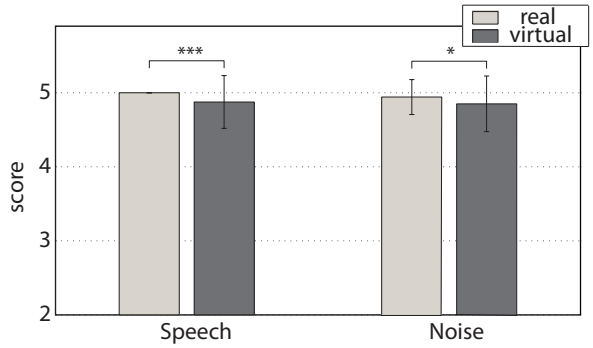


Figure 3.7: *Rating of stability for all subjects*

3.4.3.3. Classification of the Sound Source

The task of the test subjects was to detect real from virtual sources. We see that as an ultimate validation of our simulation algorithm. A perfect algorithm would result in undistinguishable sound reproductions. However, differences in color, in power or the feeling of presence captured by the whole body caused by the mass of air in movement produced by the loudspeakers are hardly reproducible by the open CICs. Despite these strong constraints, our system performed surprisingly well. The results of the third question are shown as a confusion matrix in Table 3.3.

classified as	presented as	
	Speech	
	real	virtual
real	92.5	35.83
virtual	7.5	64.16
	Noise	
	real	virtual
real	83.33	57.5
virtual	16.67	42.5

Table 3.3: *Confusion matrix in [%]*

The reasons for the decision "simulation", as reported by the test subjects, were combined into three categories. *Perceptual differences:* reasons for detecting the simulations were caused by poor externalization, diffuseness or a front-back uncertainty. When the signal was played in the front, the listeners often reported an internal image of the sound. Most of the time,

	Perceptual differences			Algorithm limitations			Sensor imperfectness			
	extern.	diffuseness	front-back	artifacts	color	reverb.	stability	off pos.	other	total
true negative	43 (33.6%)	15 (11.7%)	6 (4.7%)	3 (2.3%)	9 (7.0%)	5 (3.9%)	16 (12.5%)	22 (17.2%)	9 (7.0%)	128 (100%)
false negative	10 (34.5%)	4 (13.8%)	2 (6.9%)	1 (3.4%)	5 (17.2%)	2 (6.9%)	3 (10.3%)	1 (3.4%)	1 (3.4%)	29 (100%)

Table 3.4: *Reasons for classifying the signals as simulation. True negatives indicate that the simulation has been correctly identified. False negatives count real presentations wrongly perceived as virtual.*

this was only a first impression. As soon as the subject moved his head, the source was perceived and remained external. The listeners used this as a cue for identifying virtual sources. Some listeners were not sure if the source was in the front or in the back (front-back confusion) and concluded that such an uncertainty stems from the simulation. *Algorithm limitations*: the algorithms produced some artifacts or unnatural coloration differences when the subjects moved their heads. These processing errors could result in stuttering signals or very short pauses (dropouts), making the simulation easy to detect. *Sensor imperfectness:* the XSens sensor has a drift caused by large head movements [Damgrave & Lutters 2009]. The data is also not free of noise. All this results in simulated sources slightly off-position (located somewhere between two loudspeakers) and stability issues. The results from the open question are summarized in Table 3.4.

The results show that externalization was the main cue for distinguishing virtual from real presentations. 33% of correct identifications were based on externalization. Interestingly, real signals were also perceived as internal for 34% of the false negatives. The correct externalization of frontal sources remains the biggest challenge for virtual sound systems based on HRTFs. Getting the spatial cues right for sounds played at 0^o seems critical. Front-back confusions and change in the diffuseness of the played source played a more minor role. The proportions of decisions based on these cues are similar between true and false negatives (11% and 5% versus 13% and 7% respectively). 13% of true negatives are caused by algorithm limitations, which is relatively few considered the real-time constraints. However, a large proportion of approximations caused by the head-tracker allowed the correct identification of the virtual sources (30%). Investing in better sensors would turn this proportion down.

3.4.3.4. Position dependency

Apart from the dependency of the stimuli, the results depend also on the playback position. The externalization ratings as a function of the source position are shown in Fig. 3.6. The results of the positions at $90^o, 180^o, 270^o$ azimuth do not show any significant differences and are analyzed jointly. The stability rating is also not dependent on the playback position.

Again, the distinction between virtual and real presentations was mostly correct for positions played in the front (see Table 3.5). In the side, the listeners could hardly distinguish

classified as	presented as			
	real		virtual	
	front	sides	front	sides
real	85	88.5	40	49
virtual	15	11.5	60	51

Table 3.5: *Confusion matrix for the front position and the other positions, values in [%].*

both sound reproduction methods and half of the simulated sounds were perceived as real sources. With 40 % of the simulated presentations in the front perceived as coming from a loudspeaker, our system performs surprisingly well considered the cheap and limited hardware.

3.5. Conclusion

The system for virtual acoustics presented in this paper allows the creation of perceptually convincing virtual environments. It combines aspects of both direct and parametric room impulse response rendering techniques into an efficient real-time algorithm that works on standard Windows PCs. Head-movements are measured with a motion-sensor fixed on the head of the listeners . Depending on the positions and orientations of the head of the listener and the virtual sources, the virtual scene is updated with no noticeable delay.

The perceptual evaluation of the system showed that head movements are essential for a good externalization of virtual sources. For noise and speech signals, the perceptual differences between real and virtual sources were small, although significant. The correct externalization of frontal positions remains sensitive. Even for sounds played by an external loudspeaker located in the front of the listeners, internalization did occur. With head movements, the internalized sound images moved and remained out of the head.

Part II.

Perception of the auditory space with bilateral hearing instruments

4. Localization with bilateral hearing aids

4.1. Introduction

The human auditory system is constantly engaged in the identification and localization of various competing sources in complex acoustical environments. The everyday soundfield typically contains background noise, reverberance and simultaneous sound events coming from different directions. Despite the complexity of the acoustical scenes, the binaural auditory system is able to effectively separate and localize sound sources of interest. Sound localization is affected by background noise, reverberation and interfering signals among others [Good & Gilkey 1996, Lorenzi *et al.* 1999, Langendijk *et al.* 2001]. To localize sound sources the human auditory system uses mainly interaural time and level differences (ITDs and ILDs). Additionally, the spectral filtering induced by the pinna allows the identification of the elevation of the sound sources. Pinna cues are also essential to resolve front-back confusions.

Sound localization with bilateral hearing aids has been investigated in various recent studies with different device types, listening configurations, algorithms and microphone positions. Questionnaire surveys indicated clear benefits in sound localization for patients fitted with bilateral hearing aids compared to unilateral fittings for every type of device [Boymans *et al.* 2009, Noble & Gatehouse 2006]. Listening experiments carried out in the laboratory, however, indicate a degradation in localization performance caused by bilateral hearing aids compared to unaided conditions [van den Bogaert *et al.* 2011, Best *et al.* 2010, van den Bogaert *et al.* 2006, Keidser *et al.* 2006, Köbler & Rosenhall 2002, Noble & Byrne 1990]. In these studies, when hearing-impaired listeners were tested, the signals in the unaided conditions were played at equal loudness levels. The results suggest that, while hearing impaired subjects benefit from the amplification provided from the second hearing aid, the signal processing in the devices distorts essential localization cues.

Several factors are detrimental for the localization of sound sources with bilateral hearing aids. [Keidser *et al.* 2006] investigated the effect of multi-channel compression, noise reduction and directional microphones on horizontal sound localization. Their study included Behind-The-Ear (BTE), In-The-Ear (ITE) and Completely-In-the-Canal (CIC) hearing aids, considering thus microphone position effects as well. Their results showed that compression and noise reduction distorted ILDs, which led to a poorer performance. The position of the microphones of BTE hearing aids reduces pinna cues that are used to distinguish sounds from the front and the back. This has been confirmed in various studies [van den Bogaert *et al.* 2011, Best *et al.* 2010, Keidser *et al.* 2006, Köbler & Rosenhall 2002]. The use of directional microphones can reduce the number of front-back confusions [Keidser *et al.* 2006].

The experiments reported previously have been carried out in the laboratory with

different degrees of complexity but represent nevertheless artificial situations. In [Keidser *et al.* 2006], for example, one test condition included the presence of a constant interfering noise at 80° of the listeners whereas the other algorithms were evaluated in quiet. In [van den Bogaert *et al.* 2006], sound localization with bilateral hearing aids was evaluated in a moderately reverberant setting. In one condition, interfering multitalker babble noise was played at defined positions at the sides of the listeners. Hearing aids need to be evaluated in acoustical environments in which they are commonly used, because noise suppression algorithms affect auditory cues differently in noisy environments, depending on the type, the level and the position of the noise. Reflections might diminish the effectiveness of beamforming techniques as well.

Virtual acoustics can be used to evaluate hearing aid algorithms in more realistic environments. It is a relatively simple and convenient method for reproducing virtual spaces. This technique combines Head-Related Transfer Functions (HRTFs) and room simulations and theoretically allows the reproduction of any sound field at the eardrums of the listener [Moeller 1992]. The use of virtual acoustics enables the evaluation of existing hearing aid algorithms and research prototypes in the most diverse and relevant listening environments. The hearing aids can be implemented offline, which allows the evaluation of the most advanced algorithms. The realism of sounds generated with virtual acoustics and its impact on sound localization have been investigated in numerous studies. It has been shown that virtual sound sources can be localized as accurately as real sources when individual HRTFs are used [Bronkhorst 1995, Wightman & Kistler 1989].

In this study, four different scenes in diffuse background noise and three hearing aid algorithms were implemented. The scenes were generated using individual HRTFs and played via speakers located in the ear canals of the test subjects. The virtual environment was simulated using the ROOMSIM software [Schimmel *et al.* 2009]. The simulator uses an image source model to simulate early reflections. This is combined with a stochastic process that models late reflections. A similar room simulation procedure was used by [Rychtarikova *et al.* 2009b]. In their study, sound localization and speech intelligibility have been compared between a real and a simulated playback room. In the latter condition the Binaural Room Impulse Responses (BRIRs) were generated using HRTFs measured on an artificial head. Their results show an increase in front-back confusions for the virtual condition. It is possible that the use of non-individualized HRTFs and the impossibility to make head movements partly increased the rate of errors. No change in speech intelligibility was noticed between the two reproduction methods.

In background noise or in the presence of competing interference, sound localization degrades with decreasing signal-to-noise ratio (SNR) [Lorenzi *et al.* 1999, Good & Gilkey 1996] or when the interferer is located close to the target signal [Langendijk *et al.* 2001]. In these conditions, front-back confusions and the perceived elevation of the source are most affected by the interference. Front-back confusions however can be resolved by head movements, as shown by [Wightman & Kistler 1999] and [Wallach 1940]. Using slight head movements, the listener can resolve ambiguities in the horizontal cues and differentiate a sound in the back from the front and vice-versa.

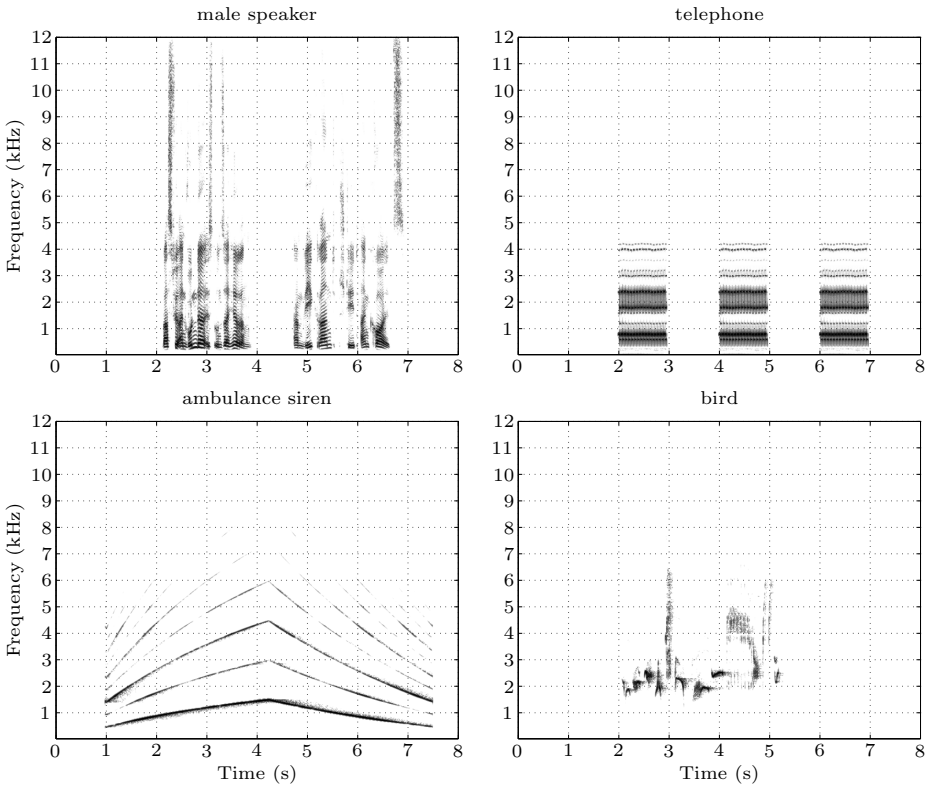


Figure 4.1: *Spectrograms of the target signals used in the localization experiment. Four scenes were implemented: a cafeteria, an office, a street and a forest. The listeners had to localize a male speaker, a phone, an ambulance siren and a bird, respectively.*

In the first experiment presented in this study, the virtual playback system was evaluated. The scenes were either played through a ring of loudspeakers located in a real room or reproduced virtually. Sound localization was then compared between the real and the virtual playback rooms. In the second experiment, the usefulness of the system for hearing aid testing was evaluated. Three standard BTE hearing aid algorithms were tested, namely an omnidirectional microphone, a cardioid-shaped beamformer and a noise canceler.

4.2. Method

4.2.1. Reference conditions

Four different scenes were selected based on their everyday relevance. The experiment required the localization of four different test signals that favor different ranges of localization cues. The four scenes are:

1. a man speaking in a crowded cafeteria.

2. a phone ringing in a busy office.
3. an ambulance siren on a busy street.
4. a bird singing in a windy forest.

The spectrograms of the four target signals are shown in fig. 4.1.

The room where the localization experiments were carried out was an acoustically treated shoebox-type room with octave-band reverberation times (T_{60}) shown in Table 4.1. The room was 6.53 meters long, 5.72 wide and 2.34 high. The receiver was set at position (3.69, 2.85, 1.15) facing the long wall. The sounds were played through a loudspeaker ring centered on the receiver position at a distance of 1.5 meters with an angular spacing of 30°.

The background noise consisted of single channel recordings of ambient sounds. The signals were sampled in twelve segments of eight seconds each. The starting points of the segments were chosen randomly along the initial sound signal. The twelve signals were then played simultaneously over the loudspeaker ring, creating a diffuse soundfield around the listener. All recordings were done using omnidirectional microphones. Target and noise were recorded separately.

For the four scenes the signal-to-noise ratio was set to 3 dB SNR based on their rms values. This SNR was chosen as containing sufficient noise for the hearing aid algorithms to work properly while maintaining good localization performance. The level of the background noise was set to 60 dB at the center of the loudspeaker ring.

In experiment I, three conditions were tested. In the first condition, the scenes were played through the loudspeaker ring in the real room. The test subjects listened with their "own ears". This is the absolute reference condition and is referred to as *ls_open*. The second condition (*sim*) evaluates the system for virtual acoustics. The playback room was simulated and the sound was played through small speakers located in the ear canals. The HRTFs used for the simulations were measured using the same devices. Due to their size, this ear canal speaker-microphone system might modify monaural spectral cues and influence negatively sound localization. Therefore, we included the condition in which the scenes are played by the external loudspeakers while the test subjects wear passive speakers (*ls_cic* condition). The ear canal transducers are described in details in section 4.2.3. In experiment II, the conditions tested are called *omni, beam* and *NC* for the omnidirectional, beamformer and noise canceler algorithms respectively. A description of the algorithms is given in the following section.

The corresponding spectrograms of the four target signals are shown in Fig. 4.1.

The room where the localization experiments were carried out was an acoustically treated shoebox-type room with octave-band reverberation times (T_{60}) shown in Table 4.1. The room was 6.53 meters long, 5.72 wide and 2.34 high. The receiver was set at position (3.69, 2.85, 1.15) facing the long wall. The sounds were played through a loudspeaker ring centered on the receiver position at a distance of 1.5 meters with an angular spacing of 30°.

The background noise consisted of single channel recordings of ambient sounds. The signals were sampled in twelve segments of eight seconds each. The starting points of the segments were chosen randomly along the initial sound signal. The twelve signals were then

played simultaneously over the loudspeaker ring, creating a diffuse soundfield around the listener. All recordings were done using omnidirectional microphones. Target and noise were recorded separately.

For the four scenes the SNR was set to 3 dB based on their rms values. This SNR was considered as containing sufficient noise for the hearing aid algorithms to work properly while maintaining good localization performance. The level of the background noise was set to 60 dB at the center of the loudspeaker ring.

In experiment I, three conditions were tested. In the first condition, the scenes were played through the loudspeaker ring in the real room. The test subjects listened with their "own ears". This is the absolute reference condition and is referred to as *ls_open*. The second condition (*sim*) evaluates the system for virtual acoustics. The playback room was simulated and the sound was played through small speakers located in the ear canals. The HRTFs used for the simulations were measured using the same devices. Due to their size, this ear canal speaker-microphone system might modify monaural spectral cues and influence negatively sound localization. Therefore, the condition was included in which the scenes were played by the external loudspeakers while the test subjects wore passive speakers (*ls_cic* condition). The ear canal transducers are described in detail in section 4.2.3. In experiment II, the conditions tested are called *omni, beam* and *NC* for the omnidirectional, beamformer and noise canceler algorithms, respectively. A description of the algorithms is given in the following section.

4.2.2. Hearing aid algorithms

The first implemented algorithm was the omnidirectional microphone configuration. In this case, the scenes were simulated using the front microphones of the BTEs only. No processing was done by the hearing aids. This condition investigated the effect of the microphone position on sound localization.

The second algorithm was a first order differential static beamformer. It has a cardioid directional characteristic and reduces sound coming from 180°. The directivity pattern was obtained by delaying the signal of the rear microphone. The frequency-dependent phase shifts depend on the distance between the front and back microphones and on the individual HRTF characteristics. The differential processing of the algorithm introduces a highpass behavior. A lowpass filter compensates for this effect [Hamacher *et al.* 2005].

The noise canceler was a Wiener filter type implementation. The incoming signal was divided into frequency bands. For each subband, the power spectra of the noise and of the speech were estimated. Subbands with high noise, i.e. low SNR, were attenuated whereas subbands with high SNR were unchanged. The SNR estimator is based on the assumption that the noise signal is relatively stationary, whereas the target is more heavily modulated [Hamacher *et al.* 2005].

Since both monaural algorithms modify level and phase independently in each hearing aid on the left and right side, ITDs and ILDs will potentially be modified. The noise canceler, however, does not change ITDs. Both algorithms were implemented on a Simulink platform

and all the processing was done offline, prior to the first test session.

4.2.3. Virtual sound reproduction

The sound recording and playback device consists of a customarily designed pair of miniature microphone-speaker systems located inside a subject's ear canal. They are mounted on an open shell of CIC hearing aids. The devices were manufactured individually for every test subject prior to the experiment. The choice of the open CIC system over headphones was due to the following reasons: first, the ear canal is open during playback. This improves the reproduced spatial image and reduces the effect of sound internalization [Kim & Choi 2005]. Second, the system always stays at the same location in the ear canals. The system therefore does not need to be calibrated at each utilization. Finally, being an open system, it allows a direct comparison between loudspeaker and simulated playbacks.

4.2.3.1. HRTFs measurements

The HRTFs were measured in a low reverberant sound-treated room using the maximum-length sequence (MLS) technique [Rife & Vanderkooy 1989]. They were recorded using the microphone of the open CIC systems. Reflections were removed from the HRTFs by trimming the impulse responses 4 ms after the first peak. The MLS signals were played at 70 dB SPL. The sequence was sampled at 44.1 kHz and lasted 6 seconds. The recordings were done using the same loudspeaker arrangement as described in section 4.2.1. The resolution of the HRTFs was thus 30°. HRTFs were measured for each participant at the beginning of the first test session.

To complete the set of measured positions, the recorded HRTFs were merged into a set of anechoic KEMAR HRTFs [Gardner & Martin 1994]. The KEMAR data set consists of HRTFs recorded on dummy head for 710 positions, ranging from elevation angle $-40°$ to $90°$ with a minimal azimuthal separation of $5°$. The direct sound component of the simulated BRIRs was always composed of the individual recorded HRTFs. The KEMAR HRTFs were exclusively used for simulating reflections where no measured transfer function was available. The generation of the BRIRs is described in details in section 4.2.3.4.

4.2.3.2. BTE HRTFs interpolation

The set of BTE Head-Related Transfer Functions (BRTFs) was recorded by a pair of standard BTE hearing aids each with two microphones at 12 mm distance. They were measured in the same room as the HRTFs and using the same procedure. The set of BRTFs was interpolated to a collection of transfer functions of the same format as the KEMAR HRTFs, covering the same positions. This was done because the algorithms are very sensitive to phase and amplitude differences between the BRTFs of the front and rear microphones. The combination of the BRTFs with unprocessed KEMAR data would reintroduce absent pinna cues as well.

The interpolation of BRTFs was carried out after time-alignment of the transfer functions. It has been shown that the performance of interpolation in the time or frequency domain can

Table 4.1: *Octave band reverberation times of the measured and the simulated rooms in [ms].*

frequency [Hz]	125	250	500	1000	2000	4000	8000
$T_{60}meas$ [ms]	230	270	270	210	230	300	300
$T_{60}sim$ [ms]	229	271	273	213	229	304	331

be improved by compensating HRTFs prior to interpolation according to the time of arrival of sound [Matsumoto *et al.* 2004b]. That is, the HRTFs were time aligned and interpolation was carried out on the time-aligned HRTFs. In order to achieve sub-sample precision in the time alignment, the time of arrival itself was also interpolated. For positions in the horizontal plane, the BRTFs were linearly interpolated after time-alignment by a factor of six giving a resolution of 5°.

For the transfer functions corresponding to positions of different elevations, the delays to the front and back microphones were obtained using the spherical-head model described in [Duda & Martens 1998]. This procedure ensured that the delays between the front and back microphones are realistic. The amplitudes were obtained by interpolating the measured BRTFs at the corresponding azimuths in the horizontal plane. The interpolated BRTFs were used only for simulating reflections.

4.2.3.3. HRTF and BRTF calibration

The HRTFs and BRTFs were measured at different positions at the ears. This induces coloration differences that need to be compensated before playback. The equalization of the transfer functions was done using the diffuse calibration method described by [Moeller 1992] (sec. 5.2, p. 197). According to this technique, the transfer functions were averaged across all measured positions. The transfer functions were then divided by the average filter of the measured positions and multiplied by the average filter of the playback positions. This removed effectively the coloration differences between two transfer functions.

4.2.3.4. Room modeling and simulation

The virtual room was a simulation of the room described in section 4.2.1. It was modeled with the ROOMSIM software. The surface absorption parameters of the ROOMSIM simulator were set to fit reverberation times measured in this playback room. The surface diffusity parameters of the ROOMSIM software were adjusted in order to match the level and the diffusity of the reflections and to minimize the perceptual differences between the simulated and measured impulse responses.The direct sound component of the BRIRs was composed of the individual recorded HRTFs or BRTFs. In this setup, most of the reflections were simulated using KEMAR HRTFs or interpolated BRTFs.

The directivity of the loudspeakers was modeled as a three-dimensional cardioid, pointing towards the receiver. This implies that most of the reflective energy came from the floor, the ceiling and the facing walls.

4.2.4. Test procedure

For each test condition, the test subject was asked to localize the target sound source (i.e. the male speaker, the phone, the ambulance siren, the bird) in the situation-specific background noise. Every test condition started with an orientation session, in which the scene was presented to the test subject. In this training round, all the twelve positions were played one after the other starting from the front and moving counter-clockwise. The subject could follow the position of the sources on a touch screen located in front of him. The diffuse background noise was played continuously. This was followed by a second training session, where every target position was presented once. Before the actual test run, the test subjects had to point out on the screen the position where they heard the sound coming from. Feedback was provided. Every position was presented twice in random order, resulting in 24 stimuli. The test subject had to indicate the position of the target source on the touch screen. No feedback was provided. The touch screen symbolically represented the test scene (i.e. cafeteria, office, street, forest) with twelve buttons arranged around a schematical listener. The subjects were instructed not to move their head during the experiment. A typical test run lasted approximatively 10 minutes.

The twelve test conditions of the first experiment (4 scenes $\times$ 3 playback modes) were divided in three blocks of four. The eight conditions where the subjects wore the open CIC devices where randomly mixed. The four other conditions were presented in one block, in random order. This made it more comfortable for the test subject, as the open CICs did not need to be repeatedly inserted and removed between two successive tests. The three blocks were presented in random order. After one block was completed (approx. 40 min), the subjects took a break. The test subjects who completed the first experiment, returned on another day for the second experiment. The twelve test conditions (4 scenes $\times$ 3 hearing aid algorithms) were randomly mixed in three blocks of four. The second experiment followed the same test protocol as the first one.

At the beginning of the experiment, the test subject was asked to match the level of the simulation to the level of the loudspeaker presentation. To do this, the listener could switch between loudspeaker presentation and simulation to compare both loudness levels. He could increase and decrease the level of the simulation in steps of 1 dB until it matched the level of the external presentation.

4.2.5. Test subjects

Twelve normal-hearing subjects took part in the experiment (9 males, 3 females, age 35 $\pm$7 years). All subjects were checked to have hearing thresholds lower than 20 dB across all frequencies.

4.2.6. Data analysis

0^o was defined here as the position directly in front of the listener, 90^o as the position to the left of the listener, and 270^o as the position to the right of the listener. The localization per-

formance was evaluated in two different ways. The accuracy of the directional localization was measured using the angular root-mean square (rms) error. As another indicator of the quality of the simulation, the amount of front-back confusions (fb) was considered. Front-back confusions occur when a sound presented in the front is heard in the back and vice-versa. Those two phenomena represent different types of errors and were analyzed separately. Furthermore, the standard angular rms error is particularly sensitive to front-back confusions. Such confusions cause large errors for positions where the directional information was perceived and reported correctly. To remove this effect, the front-back confusions were resolved prior to measuring the directional error, which has commonly been done in localization experiments [Langendijk *et al.* 2001]. The angular rms error rms_θ is defined for each position as followsh:

$$\mathrm{rms}_\theta = \sqrt{\frac{\sum_{i=1}^{N} \left(\arcsin\left(\sin x_\theta\right) - \arcsin\left(\sin y_{\theta,i}\right)\right)^2}{N}} \tag{4.1}$$

where x_θ is the position played at angle θ and $y_{\theta,i}$ the response given by the test subject at test iteration i. N is the total number of repetition. Eq. 4.1 implies that for a sound source played at 30°, 30° and 150° are considered to be correct answers. An average rms error taken over all played positions characterizes a subject's directional performance for a given test condition.

The amount of front-back confusions was evaluated as a percentage of occurrence over all possible confusions. Positions played at 90° and 270°, for which front-back confusions are not defined, were ignored. Sounds incorrectly located at 90° and 270° were not considered as confusions. These corrections result in a chance level of 41.66%.

4.3. Results

4.3.1. Experiment I: Evaluation of the virtual acoustics system

Fig. 4.2 shows the rms_θ averaged across all test subjects for every test condition. The rms_θ varies considerably across position and across scene. In all scenes except the cafeteria, the sound was accurately localized in the front, but rather poorly on the sides or in the back. The same pattern appears for presentation over loudspeaker with or without CICs (*ls_ open* and *ls_ cic* conditions in the upper and middle panels) and for the simulation (*sim*, in the bottom panel). The four scenes were not perceived as equally difficult. The male speaker was easily localized whereas the bird's position was frequently misjudged.

The upper panel of Fig. 4.3 shows the rms error, averaged across test subjects, for every test condition along with one standard deviation. The test subjects performed differently in the four different scenes. The overall results for each test condition are shown in Table 4.2.

Significant differences between the rms error and the amount of front-back confusions for the four scenes and the three reproduction methods were examined using a one-way analysis of variance. Significance was set at $p < 0.05$. No significant difference in terms of rms error between the three reproduction methods was found for the *office, street* and *forest* scenes

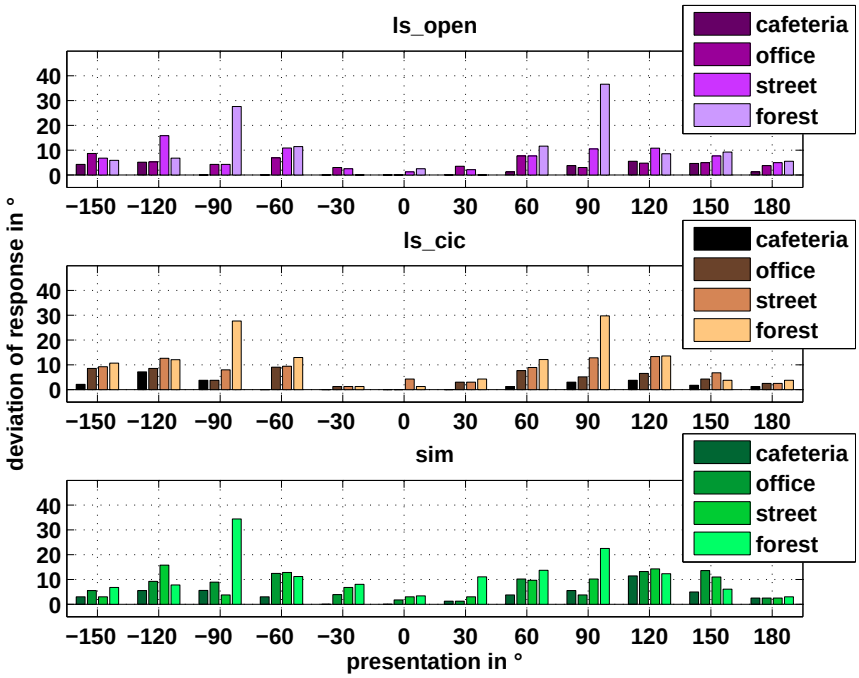


Figure 4.2: *Mean angular rms error rms_θ for the different scenes for the three sound reproduction methods. ls_open denotes loudspeaker playback with open ear canal (the natural listening condition), ls_cic stands for loudspeaker playback with the open CICs in the ears and sim is for fully simulated environments with sound playback through the open CICs.*

Table 4.2: *Mean results and standard deviations for all the scenes tested. The last column shows performance averaged across scenes. f-b denotes front-back confusions in* [%].

	cafeteria			office			street			forest			all		
rms(°)	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim
mean	2.2	2.0	3.9	4.7	5.0	7.2	7.1	7.7	8.0	10.5	11.1	11.7	9.1	9.3	10.4
std	2.3	2.2	2.2	3.7	3.8	3.7	3.3	3.4	2.8	2.5	2.9	4.4	2.5	2.6	2.7
f-b (%)															
mean	1.0	2.3	7.5	11.7	12.1	10.8	6.5	10.8	13.1	10.4	7.7	10.4	7.4	8.2	10.5
std	2.0	5.3	6.8	10.5	7.4	7.1	7.2	9.6	3.9	7.7	5.8	8.0	5.1	6.0	4.2

($p > 0.11$).

The localization was only significantly worse in the simulated *cafeteria* condition (rms, fb: $p = 0.05$). The average rms error of the *sim* condition in this scene was, however, very small (3.9°). For the other sound reproduction methods, the localization of the target speaker was nearly perfect with a directional error of at most 2.2° and 2.3% of front-back confusions. The passive open CICs in the ear canal did not impair localization performance ($p \geq 0.23$).

The amount of front-back confusions varied greatly with the test subjects. As a result,

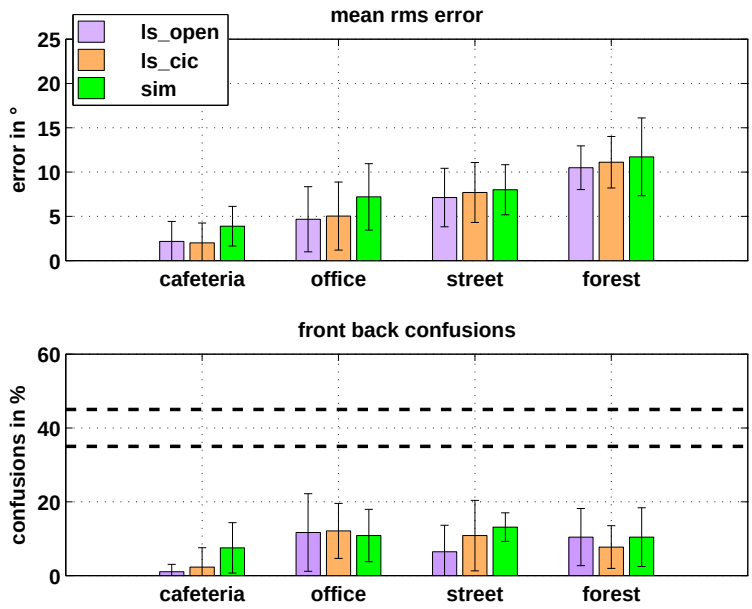


Figure 4.3: *Mean rms error (above) and percentage of front/back confusions (below) for each reproduction methods for the different scenes. The error bars show one standard deviation. Chance level, along with 95% confidence interval is plotted in dashed.*

the standard deviations were very large. The bottom panel of Fig. 4.3 shows the percentage of confusions for the different scenes and reproduction methods. The dashed lines show the chance level along with the 95% confidence interval. Results falling in this interval can be considered to follow with a 95% certainty a random guessing strategy.

For the *office* and *forest* scenes, the amount of front-back confusions was similar for the three reproduction methods. In these conditions, the simulations did not affect localization ability. In the *cafeteria* scene, the simulations were significantly worse than the *ls_open* and *ls_cic* conditions ($p \leq 0.05$). 7.5% of the signals were incorrectly localized in the front or in the back, which was significantly more than 1.0% and 2.3% for the *ls_open* and *ls_cic* conditions respectively. In the cafeteria condition, the target signal was the most broadband of the stimuli presented, containing low frequency components. At low frequencies, the human auditory system is sensitive to ITDs as small as 10μs [Hershkowitz & Durlach 1969]. At a sampling rate of 44.1 kHz, this corresponds to half of the sample interval. Measuring ITDs with this precision is difficult. This would explain why the *cafeteria* case was the only scene where the *sim* condition yielded significantly worse performance in terms of directional and front-back errors, even though it was perceived as the most easy by the test sujects. Although the subjects were told to keep their head in a fixed position, unintentional head movements could have helped to resolve the front-back confusions when the sound was played through the loudspeakers. The virtual system was not set up to respond to head movements in this experiment.

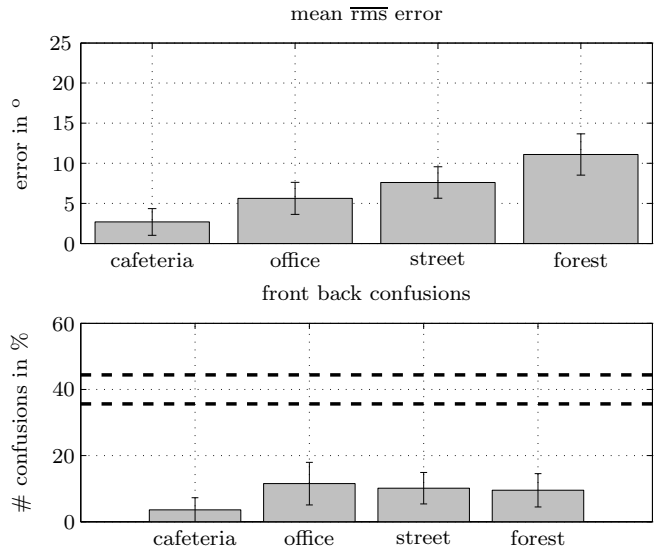


Figure 4.4: *Mean total directional $\overline{rms}$ error (above) and number of front-back confusions in % for the different scenes. The data for the three reproduction methods were pooled together.*

The four different scenes were not perceived as equally difficult (see Fig. 4.4). This was desired as the aim of the second experiment was to explore the weaknesses of different hearing aid algorithms. Scenes with different characteristics and degrees of difficulty permit to better rate and evaluate the hearing devices. The results clearly show a change in rms error. Averaged

across the three reproduction methods, the rms error was 2.1° for the *cafeteria*, 4.5° for the *office*, 5.9° for the *street* and 9.7° for the *forest* condition. The rms error differences between the *cafeteria* and *forest* scenes and the other environments were statistically significant. The amount of front-back confusions was relatively similar between the different scenes, with the *cafeteria* showing slightly fewer mistakes (3.6% vs. 11.4%, 8.9% and 10.2% for the *office*, *street* and *forest* respectively).

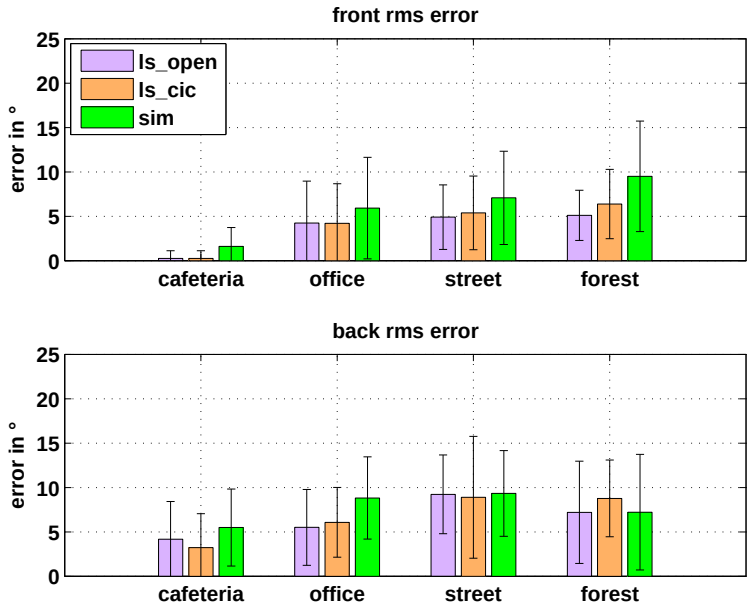


Figure 4.5: *Mean rms error for positions played at front ($|\theta| \leq 60^o$, above) and in the back ($|\theta| \geq 120^o$, below.)*

The rms error was higher in the back than in the front, as illustrated in Fig. 4.5. For signals played in the front, the statistical analysis showed again that performance in the *cafeteria* and *forest* scenes was significantly worse for the *sim* condition ($p \leq 0.05$).

The open CICs affect mostly the high frequency content of the signals, due to their small size. For signals played in the back, high frequencies are naturally attenuated by the pinna. An inaccurate reproduction of high frequencies has therefore less effect than for signals played from the front. This could be an explanation for the difference in localization performance that can be seen in the front, but not in the back. No significant differences were found for back positions.

The effect of learning on the performance of the test subjects was further examined. No significant difference was found between the test and retest sessions.

Table 4.3: *Mean results and standard deviations for the BTE conditions. The last column shows performance averaged across scenes.*

	cafeteria			office			street			forest			all		
rms(°)	omni	NC	beam	omni	NC	beam	omni	NC	beam	omni	NC	beam	omni	NC	beam
mean	7.9	8.7	10.3	8.2	8.3	7.0	11.2	11.8	12.7	15.4	16.1	16.2	13.7	14.3	14.7
std	4.0	3.9	4.0	3.9	3.7	4.7	3.1	3.2	3.2	2.8	4.2	2.8	2.4	2.4	2.0
f-b (%)															
mean	39.0	39.2	1.9	40.8	37.1	1.0	45.4	48.3	3.8	47.3	43.3	1.5	43.1	42.0	2.0
std	11.2	8.9	3.2	12.2	11.9	2.0	7.9	8.3	5.5	6.4	10.1	3.1	5.8	6.0	2.5

4.3.2. Experiment II: Evaluation of BTEs algorithms

In the second experiment, the localization task was repeated with the same subjects and the same scenes. All signals were processed offline and presented through the open CICs. Three different BTE algorithms were evaluated: the omnidirectional case (*omni*), where no processing was done by the hearing aid, the beamformer (*beam*) and the noise canceler (*NC*) conditions with active BTE processing. The results were analyzed in the same way as for experiment I. They are shown in Fig. 4.6, with the upper panel displaying the directional rms error and the lower panel the amount of front-back confusions in %. Chance level lays between the two dashed lines. As a reference, the *sim* condition as reported in section 4.3.1 is shown as well. The average errors and standard deviations for all test conditions are shown in Table 4.3.

Considering directional rms errors, the differences between BTE algorithms was not statistically significant. The amount of front-back confusions did not significantly differ between the *omni* and *NC* cases ($p \geq 0.27$). Due to the strong attenuation characteristics of the beamformer, the listeners could clearly identify sound coming from the back based on intensity cues in the *beam* condition. For this algorithm, some subjects verbally reported some front-back confusions, especially for sound being played at 0^o, but responded correctly on the response map. For all scenes but the *office* scene, subjects performed significantly worse for the algorithms compared to the virtual simulations ($p \leq 0.02$). No statistical difference in directional errors between the algorithms and the simulation was found in the *office* case.

In Fig. 4.7 and 4.8, the results from positions played in the front and in the back were analyzed separately. As expected, for the *beam* condition, the rms error was lower in the front than in the back. This is due to the greater SNR in the front than in the back; improving therefore localization in the frontal area. Performance for all scenes but the cafeteria was similar for the beamformer as compared to the reference condition ($p \geq 0.25$).

For the *omni* and *NC* conditions, the error-rate was larger in the front than in the back, both in terms of rms errors and amount of front-back confusions. In the back, directional performance was similar to the reference condition.

By separating the front-back confusions that occurred in the front from the ones in the back, a pattern emerges for the *forest* scene. It appears that the target signal was mostly localized in the front, whereas performance was close to chance for the other scenes. The reason for this can be explained by the spectral content of the target signal. It is the only

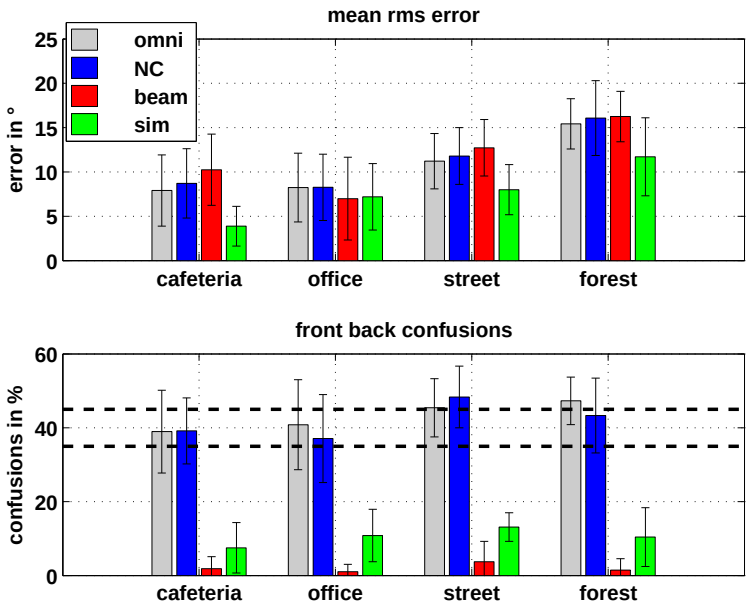


Figure 4.6: *Mean rms error (above) and percentage of front/back confusions (below) for the sim (reference, taken from Fig. 4.3), omni, NC and beam algorithms for the different scenes. The error bars show one standard deviation. Chance level along with 95% confidence interval lays between the two dashed line.*

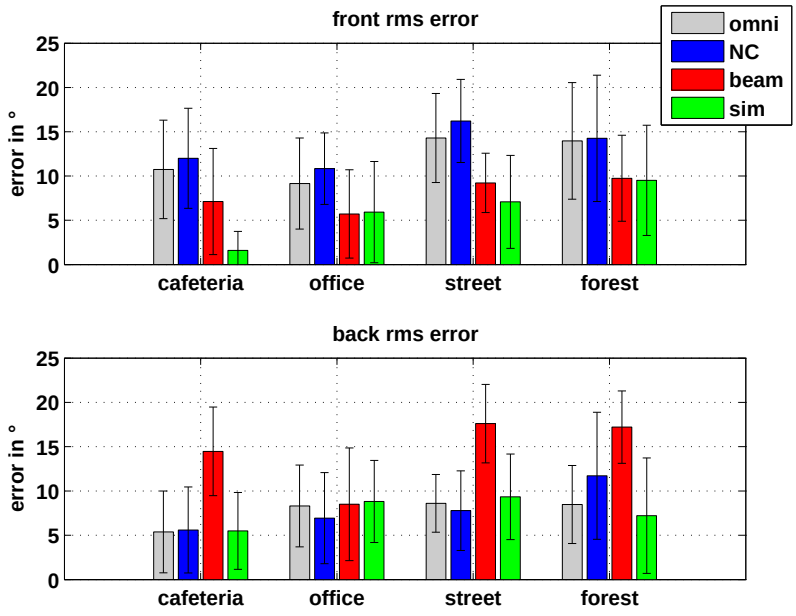


Figure 4.7: *Mean rms error for positions played at front ($|\theta| \leq 60^o$, above) and in the back ($|\theta| \geq 120^o$, below.)*

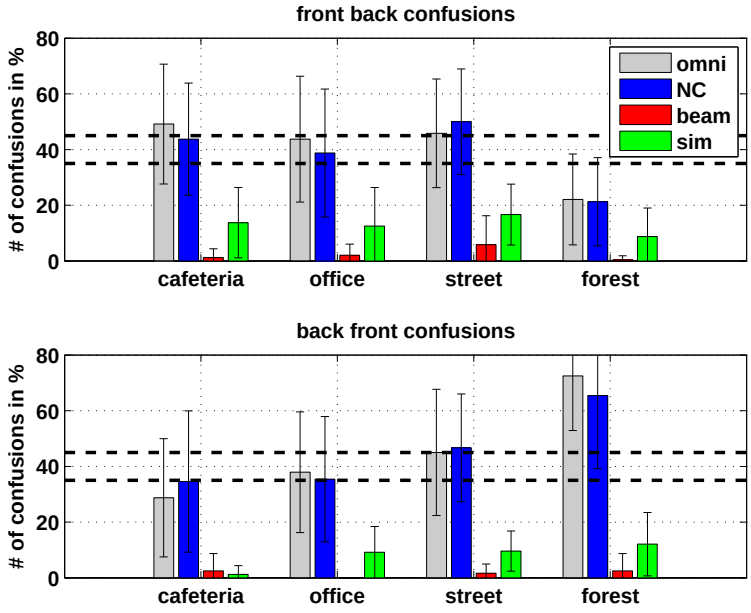
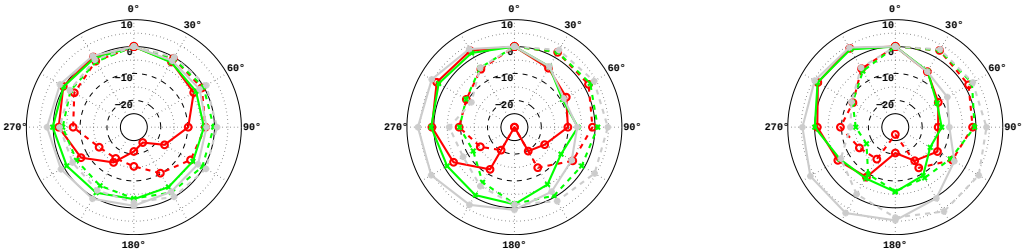


Figure 4.8: *Percentage of front-back confusions for positions played at front ($|\theta| \leq 60^o$, above) and in the back ($|\theta| \geq 120^o$, below.) Note that the y-axis has been rescaled.*

signal that is essentially composed of frequencies above 2 kHz (see figure 4.1). In Fig. 4.9, the directivity patterns of the beamformer, the HRTFs measured at the entrance of the ear canal and at the position of the BTE microphones are represented. At low frequencies, the intensity diagrams for both HRTF measurement positions are similar. For the octave-band centered at 4 kHz, the effect of the pinna-loss is clearly visible with a difference of 10 dB. The BRTFs at this frequency band were similar for the front and the back. A sound composed of high frequency is therefore heard as coming from the front.



(a) $f_c = 250$ Hz (b) $f_c = 2000$ Hz (c) $f_c = 4000$ Hz

Figure 4.9: *Directivity characteristics of HRTFs measured at the ear canal with the open CIC microphones (green), behind the ear with the BTE microphones (gray) and of the beamformer (red) implemented at three different frequency bands. The directivities of the transfer functions measured at the left ear are plotted as a solid line. For the right ear, they are drawn in dashed lines.*

4.3.3. Evaluation of hearing impaired subjects

In the framework of this project, sound localization was investigated with six hearing impaired subjects. The test subjects were selected because they suffered from a moderate and symmetric impairment. The audiograms of their hearing losses are shown in Fig. 4.10.

The cafeteria and office scenes were implemented and played over loudspeakers as the open CICs were not able to amplify the sounds enough. The street and forest conditions were not tested as they were to difficult. The test subjects were equipped with BTE hearing aids. The algorithms evaluated were the omnidirectional and beamformer microphone configurations as discussed previsouly. The results are plotted in Fig. 4.11 along with the data of the normal hearing listeners. The results are separated into rms errors and % of front-back confusions.

For all conditions, the rms errors of the hearing impaired listeners were 10^o higher than the normal hearing subjects. Due to the large variation in the performance of the hearing impaired and the few number of subjects tested, no significant difference in rms error between the omnidirectional and beamformer algorithms could be detected. The office scene appears to be slightly more difficult on the average for the omnidirectional algorithm, but this trend is not significant.

The percentage of front-back confusions is lower for the beamformer algorithm, as expected. In the office scene, the percentage of confusions was 8%, significantly higher than in

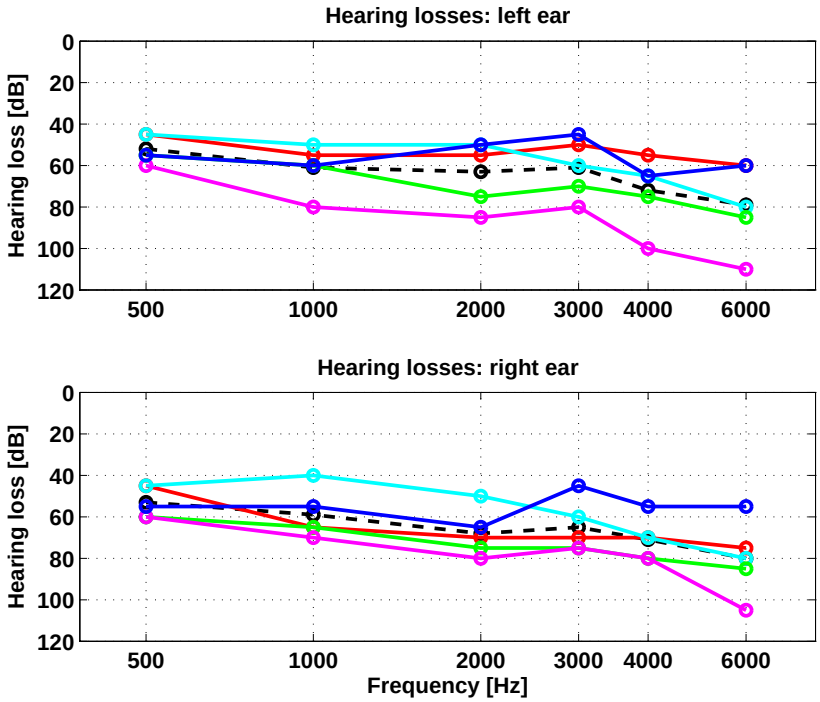


Figure 4.10: *Audiograms of the hearing impaired listeners that took part in the localization experiment.*

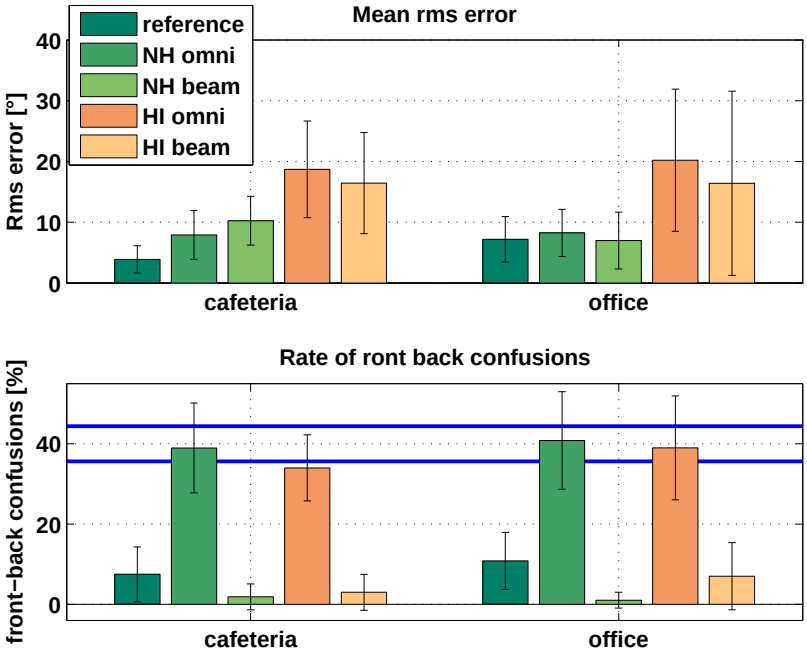


Figure 4.11: *Outcome of the localization experiments for the hearing impaired listeners (orange). The data of the normal hearing subjects for the same algorithms and the same scenes is reproduced for comparison in green. The reference (sim condition, dark green) is shown as well.*

.

the cafeteria scene. This implies that even with the attenuation provided by the beamformer, the hearing impaired listeners had some difficulties in separating sources between the front and the back. In the omnidirectional condition, the front-back confusions were at chance level.

4.4. Discussion

The aim of this study was to investigate sound localization in realistic acoustical conditions in subjects using bilateral hearing aid algorithms. For this purpose, artificial environments with background noise were reproduced using individual HRTF measurements and room simulations. The study involved various aspects of human sound localization that are discussed separately in the following sections.

4.4.1. Sound localization in noise

In this study, the listeners had to localize a sound signal in diffuse background noise. It has been shown earlier that interfering noise has an impact on sound localization. Depending on the intensity and the position of the noise relative to the target signal, localization performance is reduced [Lorenzi *et al.* 1999, Langendijk *et al.* 2001].

The influence of noise intensity on sound localization was investigated by [Lorenzi *et al.* 1999]. The task of the listeners was to localize lowpass, highpass and broadband 300 ms pulse trains played by loudspeakers placed on the horizontal plane in the frontal hemisphere. The masking noise was located at fixed positions (either at 90° (right) or 0° (front)) and its intensity was varied. Their findings showed that sound localization is affected by noise at negative SNRs only, for all tested configurations. In the present localization experiment, the signals were played at 3 dB SNR. The hearing aid algorithms further improved the SNR. According to the findings of [Lorenzi *et al.* 1999], the improvement in SNR achieved by the algorithms had no effect on localization performance, as the tested SNRs are above levels at which localization was affected. The difference in performance between localization with BTE hearing aids and the reference condition can solely be attributed to a distortion in spatial cues produced by the BTEs.

In the two localization experiments discussed in this study, the rms errors of the test subjects were lower in the frontal hemisphere than in the back. This phenomenon appears in a series of sound localization experiments [Makous & Middlebrooks 1990, Good & Gilkey 1996, Gilkey & Anderson 1995, Carlile *et al.* 1997]. In those studies, larger errors in the back than in the front for normal-hearing subjects were consistently observed for different stimuli (pulse trains, words, broadband noise) and attributed to the experimental setting of the tests. [Carlile *et al.* 1997] and [Makous & Middlebrooks 1990] evaluated localization performance using a head-pointing method, which required subjects to move their head to the direction of the sound being played. They assumed that the difference between front and back localization performance was due to the higher difficulty and time needed to move the head to the target in the back. This was different in experiments I and II, where subjects could report the rela-

tive position of the source directly on a screen in front of them, while the target was played continuously. It can be argued that the subjects had more difficulties in visualizing the exact positions of the loudspeaker in the back compared to the front, where direct visual feedback was available. This could have increased the uncertainty of their localization judgment and thus lower overall localization performance.

4.4.2. Localization of virtual sound sources

The utilization of HRTFs and room simulations for the generation of virtual acoustical environments has been discussed in previous studies. It is difficult to compare other virtual sound reproduction techniques with the system evaluated in experiment I due to the large differences in the conditions tested. In a recent study [Rychtarikova *et al.* 2009b], however, the localization of virtual sound sources was investigated in conditions similar to those of the *office* scene. Their study compared the localization of signals generated with loudspeakers versus sounds generate with HRTFs combined with room simulations or BRIRs measured in the testing room. The sounds were reproduced using headphones and the HRTFs and BRIRs were recorded on an artificial head. In one of their setups, the test subjects had to localize a telephone signal as in the *office* condition. The target signals were located either at 1 or 2.4 meters of the listeners (1.5 in our study). In the reverberant room condition and for signals played at 1 meter from the listeners, the average rms errors they obtained were 8.3°, 9.1° and 10.5° for loudspeaker playback, simulated BRIRs and measured BRIRs, respectively. For signals played at 2.4 meters distance from the test subjects, the rms errors increased to 9.9°, 11.5° and 15.1°. In the anechoic room, performance improved to 7.3° and 7.8° for loudspeaker and headphone playback respectively. In this latter condition, the telephone signal was played at 1 meter from the listener. In the *office* condition, the rms errors were 4.7° and 7.2° respectively for the *ls_ open* and *sim* conditions. These results are of the same order as in the anechoic settings in [Rychtarikova *et al.* 2009b]. The difference in performance between the two studies in reverberant settings is due to the different reverberation time ($T_{30} > 4$ s at 1 kHz in [Rychtarikova *et al.* 2009b]).

[Hawley *et al.* 1999] evaluated localization ability and speech intelligibility for a target speaker along with interfering speech from the same talker. The number of competitors varied from none to three. The evaluation was carried out both using loudspeaker playback and virtual acoustics. They evaluated positions at the front only ($-90°$ to $90°$ with steps of $30°$). Their experimental setup can be compared to the *cafeteria* condition of the present study. They measured a significant difference between real and virtual listening conditions but not between the number of competing talkers. Their rms errors were higher than in our *cafeteria* condition, being 10° for real and 14° for virtual playback (compared to 2.2° to 3.9° in experiment I). This difference can be explained by the small number of subjects doing the localization experiment (three) compared to this study (twelve). A single error on one trial results in an overall large increase in rms values. The average percentage of correct responses reported by [Hawley *et al.* 1999] was of 96 % and 83 % for both reproduction methods against 95 % and 86 % in experiment I and are therefore similar.

The system for auralizing the virtual scenarios applied static HRTFs and was therefore

not able to cope with head movements. The differences in the number of front-back confusions between loudspeaker playback and simulation in the first experiment can be attributed to this effect. Although the test subjects were instructed not to move the head, unintentional head movements may have occurred and may have been an advantage for the real versus virtual test conditions. This is especially true for the cafeteria scene, where the target signal had the largest low frequency content of all the scenes tested. This implies that interaural time differences are essential for the correct localization of the sound source. For positions played around 0° (or 180°) even small head movements can help finding the true position of the source.

4.4.3. Hearing aid localization

The bilateral hearing aid algorithms evaluated in this study had a significant impact on sound localization, although the differences in the average rms error between the omnidirectional and noise canceler conditions were rather low. This can be due to the limited number of measured positions, which might have reduced the sensitivity of the experiment. The main effect observed was an increase in front-back confusions caused by the loss of the pinna cues due to the positions of the microphones of the hearing aids. The directivity of the beamformer resolved these ambiguities. By analyzing separately the results of the front and back playback positions, it appears that the beamformer performed better in the frontal area than the other algorithms. It performed, however, much worse in the back due to reduced audibility of the target signal. In their study, [Keidser *et al.* 2006] evaluated similar algorithms. Their reference, cardioid/cardioid and max. noise reduction conditions corresponded to *omni*, *beam* and *NC*, respectively. Their findings were consistent with the results of the second localization experiment, although the test conditions were different. The first two conditions were evaluated in quiet and the noise reduction algorithm was evaluated with a constant noise source at 80° with an SNR of 7dB and the target stimulus was pink noise. They observed a slight but significant decrease in localization performance between the noise reduction algorithm and the reference condition. The cardioid microphone conditions also helped reduce front-back confusions.

Sound localization with bilateral hearing aids was examined by [van den Bogaert *et al.* 2006]. In a reverberant room (T_{60} = 0.54s), they investigated the localization of low-frequency and high-frequency noises and a telephone signal. The test subjects were normal hearing and hearing impaired subjects wearing real hearing aids. For the telephone signal, interfering noise was played at both sides of the subjects with a signal-to-noise ratio of 0 dB. The noise consisted of a multitalker babble. The test signal was played over 13 loudspeakers, situated at the frontal plane of the test subjects with an inter-speaker spacing of 15°. For the telephone signal, the normal-hearing subjects obtained an rms error of 11.8° in noise, which is higher than in the *office* condition (4.7°). The different test conditions between the two experiments could partly explain this difference. The spatial configuration of the noise sources between the two experiments differed. In [van den Bogaert *et al.* 2006], the noise source were played from two loudspeakers at both sides of the test subject whereas in our case the interfering noise was diffuse and played via

12 loudspeakers placed around the listener. For a fixed SNR, [Langendijk *et al.* 2001] showed that sound localization was more difficult when the interfering noise and the target signal were close to each other. The local SNR was lower in [van den Bogaert *et al.* 2006] than in the present experiment. By looking at the localization plots shown in their study (Fig. 5), it appears that the addition of the masker increased the errors only at positions close to $\pm 90^{o}$. For the hearing impaired, the localization worsened with and without hearing aids. Without hearing aids and with matched loudness levels, the rms error was 15.3^{o}. Without noise, the rms errors were 13.0^{o} and 16.1^{o} respectively. With the *omni* settings, the error increased to 21.3^{o}. This compares to 5.8^{o} to 9.2^{o} for the *sim* and *omni* cases (results for the frontal hemispheres only). In a subsequent experiment, group of hearing impaired subjects was tested with the same hearing aid algorithms. They obtained an rms error of 20^{o} for the cafeteria scene and the omnidirectional algorithm. This rms error value is similar to Van den Bogaert's results.

In open hearing aid fittings, the acoustic wave bypasses the hearing aid and reaches the eardrum before hearing aid processing and playback. This direct acoustic path can provide undistorted localization cues to the hearing aid users and improve sound localization performance, provided enough residual hearing remains [Byrne *et al.* 1996]. Furthermore, when the processing delay of the hearing aid is higher than 2 ms, the precedence effect ensures that the perceived position of the sound source is defined by the original acoustical wave [Litovsky *et al.* 1999]. In the present experiments, this direct acoustic path has not been simulated as the study focused on the effects of the hearing aid algorithms on sound localization. For subsequent studies with hearing impaired listeners, this aspect must be considered so that the testing conditions are more realistic and closer to the hearing aid users daily experience.

4.5. Conclusion

In agreement with previous research, the outcomes of the localization experiments carried out in this study, suggest that by combining HRTFs with room simulations one can create acoustical environments that sound convincing and in which localization ability is preserved. A significant increase in front-back confusions with virtual playback was noticed only for one of the four scenes simulated. This is a common problem in virtual sound localization experiments and can be related to the inability of our sound reproduction system to cope with head movements. This could be improved by combining a head motion sensor with the system for virtual acoustics.

The localization experiments carried out in this study took place in noisy and realistic scenes in which hearing aids traditionally operate. The results are consistent with findings from earlier experiments that were carried out in the laboratory in much simpler acoustical conditions. In particular, the experiments presented here showed that bilateral hearing aids distort the spatial perception of sound. However, the algorithms tested represent only a small sample of what is available on the hearing aid market today. Specifically, new binaural algorithms that were designed to reproduce correctly the interaural cues have been developed. The real benefits of these algorithms need to be evaluated in realistic conditions. Moreover,

other dimensions of spatial auditory perception such as the internalization, the perceived distance or the diffuseness of sound sources need to be investigated. The previously described setup allows the evaluation of these aspects.

The system for virtual acoustics is capable of reproducing environments that are more dynamic and closer to the real-world. In such environments, sound sources move along defined trajectories in space and in time. The behavior of adaptive algorithms is strongly linked to the environment in which they are used. Virtual acoustics could help to understand how spatial perception is affected by these algorithms and speed up the development of new binaural hearing aid prototypes.

5. Localization with bilateral Cochlear Implants: influence of head movements

5.1. Introduction

Human listeners tend to move their heads towards the source of interest in acoustically challenging environments in order to optimize the amount of useful information. Source and head movements enable the human auditory system to exploit the variation of binaural information for a better localization of the source of interest. Additionally, the combination of auditory information and visual cues, such as lip reading, significantly improves speech understanding [Munhall *et al.* 2004, Grant 2001]. Furthermore, in multi-talker environments, knowledge about the position of the target source results in a strong increase in speech recognition [G. Jr. Kidd *et al.* 2005]. This implies that the correct localization of sound sources can provide essential gains to for cochlear implant recipients for the comprehension of speech. However, speech intelligibility tests are generally carried out in static laboratory settings and therefore underestimate the real benefits of the auditory prostheses.

Previous studies showed [Bronkhorst 1995, Mackensen 2004, S.Perrett & Noble 1997, Wallach 1940, Wenzel *et al.* 1993, Wightman & Kistler 1999] that sound source and head movements help normal-hearing listeners to distinguish between sounds coming from front and rear positions. In such situations, the interaural time and level differences are similar in the front and in the back for various positions in space, within the "cone of confusion" (see Fig. 2.6, [Wightman & Kistler 1999]). In this case the only sources of information available for making this distinction are pinna and visual cues. Disregarding vision, the spectral filtering introduced by the shape of the outer ear affects sound differently whether it is played in the front or in the back. Due to the position of the microphones and the limited frequencies reproduced by the cochlear implant, pinna cues are of little use for CI users. Changes in interaural information caused by moving sources alone however can only be informative when the listener has prior knowledge about the direction of motion of the source [Wenzel *et al.* 1993]. In [S.Perrett & Noble 1997] head movements offered significant advantage in resolving front-back confusions for a low-pass filtered noise of 500 ms. This study suggests that even for relatively short signals, normal hearing subjects can use head movements and significantly reduce the number of front-back errors.

To date, sound localization with bilateral cochlear implants (CIs) has been investigated in numerous studies with a broad range of stimuli but the ability of CI users to use head movements for localization has never been explored. Depending on the experimental conditions, their ability to localize sounds with their clinical processors can be relatively high, reaching an average angular RMS error of 9.8° in [van Hoesel 2004], for example. In the same

study, the sensitivity of bilateral CI users to Interaural Time and Level differences (ITDs and ILDs) was tested. In a lateralization experiment, the best test subjects were capable of discriminating ILDs of less than 1 dB and ITDs of around 100-160 μs. In other subjects, just-noticeable-differences in ITD of 1 ms were observed, which is greater than the range of naturally occurring ITDs. The sensitivity to interaural differences is further affected by the position of the electrode arrays in each cochlea. Long [Long 2000] for example observed that a displacement of an electrode of 2 mm could influence ITD detection strongly, while ILDs were less affected.

While it is established that bilaterally implanted CI recipients can detect small ITDs and ILDs and perceive shifts in positions of the source in their head, the question remains how this translates to sound localization. To answer this question, Seeber and Fastl [Seeber & Fastl 2008] manipulated Head-Related Transfer Functions (HRTFs) measured at the microphones of the speech processors of bilateral CI users. Their findings show that ITDs play a secondary role in sound localization for low-pass, high-pass and broadband noises. A significant dominance of ILDs for the same stimuli was observed. This contrasts with normal hearing subjects were ITDs are more important for lowpass and broadband stimuli [Macpherson & Middlebrooks 2002].

In summary, the low resolution of binaural information available to the CI users and the absence of pinna cues due to the position of the microphones make them very prone to front-back confusions. The fact that they essentially rely on ILDs for localization judgments makes them less sensitive to small head movements. To evaluate the ability of bilateral CI subjects to take advantage of head movements in sound localization, an experiment has been designed whereby the test subjects had to localize target speech signals of different lengths. The working hypothesis was that longer duration target signals would benefit from head movements and thus the variation in interaural information would be larger than for short duration stimuli. A similar experimental paradigm has been used by Perrett and Noble [S.Perrett & Noble 1997] where the test subjects had to localize white noises of 0.5 and 3 second duration with and without head movements. To get closer to real-world situations, a diffuse background noise was played through an array of loudspeakers around the test subjects. To record and monitor the head movements, we used a head motion tracker (Xsens, MTx, [Technologies 2010]). A similar head tracker has been evaluated by Kerber and Seeber in [Kerber & Seeber 2009] and no interference between the cochlear implant and the motion sensor was observed.

5.2. Methods

The task of the test subjects was to localize a target male speaker in a diffuse background cafeteria noise. The speech material was taken from the OLSA test database [Wagener & Brand 2005]. Three different signal lengths were used, consisting of single names, single sentences and two sentences for Short, Middle and Long durations (503 ms, 2.18 s and 4.45 s respectively.) For each signal length, six different signals were chosen and presented in random order. This was done in order to prevent the listener from tuning into a particular stimulus. For the singla name, the selected material was: {*Stefan, Doris, Nina, Wolfgang,*

Thomas, Tanja}The speech signals were selected to have the most uniform length and loudness as possible across test conditions. The cafeteria noise was played incoherently from twelve loudspeakers located at 1.5 meters around the listener. The level of the background noise was set to 60 dB SPL and measured with a sound level meter at the center of the loudspeaker ring. The level of the target speech was set to the level of the noise based on their respective RMS values. To ensure that the listeners could not identify a loudspeaker based on a specific coloration or intensity difference, the level of the target was roved by 2 dB between successive presentations.

A typical test session consisted of two blocks of three conditions. In each block the test subject was instructed either to keep his/her head still or to move the head in the horizontal plane. In each condition, speech signals of a given duration were used. Testing in each condition consisted of an initial training phase with feedback where every position was played once in random order. The training was followed by a test run, in which every position was played twice. At the beginning of every block an orientation sequence was played, in which the signal of middle length was played from position to position, starting from the front and moving counter-clockwise. During this phase the listeners were asked to pay close attention to the position of the source. The order of the blocks and test conditions were randomized between subjects. A typical test block lasted 30 minutes. It was followed by a break and then the second test block. In total, the experiment required two sessions of one and a half hours each on different days (test-retest).

5.2.1. Test subjects

Eleven normal hearing subjects (age 36 ± 10.5 years) and seven bilateral CI users (age 56 ± 9.7 years) participated in the experiment. TThe hearing loss of the normal hearing subjects was measured by standard clinical audiometry and did not exceed 20 dB hearing loss across all frequencies. Only CI subjects with bilateral implants were included in the study. The CI recipients used a variety of speech processors which are listed in Table 5.1 together with their age, gender and date of implantation. The study was conducted in accordance with the guidelines established in the Declaration of Helsinki. Ethics committee approval was obtained.

Subject		Left ear		Right ear	
ID	Age/gender	CI proc. type	Date of Implantation	CI proc. type	Date of Implantation
BT	55/f	Freedom	Oct-02	CP810	Sep-07
HR	65/m	CP810	Feb-04	Esprit-3G	Feb-05
MA	56/m	Freedom	Nov-07	Freedom	Apr-07
NT	36/m	Freedom	Jun-02	Freedom	May-01
WK	56/f	SPRINT	Feb-00	CP810	Aug-10
WKu	64/m	CP810	Mar-10	Freedom	Dec-90
WM	62/f	Freedom	Oct01	Esprit-3G	Mar-02

Table 5.1: *Right and left Cochlear Implants along with time of implantation.*

5.2.2. Data analysis

5.2.2.1. Analysis of localization performance

The localization errors are divided into two categories: front-back confusions and directional RMS errors. Front-back confusions are caused by the similarity of interaural information along the cone of confusion. Moreover, the CI users cannot use the directivity and the spectral filtering introduced by the shape of the pinna, due to the position of the microphones of the speech processor. Azimuthal uncertainty however, measured here by the RMS error, is caused by the inability to resolve adjacent positions. It is related to the distinction of sets of interaural information corresponding to neighboring locations. Head movements are assumed to help reducing front-back confusions and increase azimuthal accuracy depending on the speed and duration of movement. Both types of errors are not truly independent. A front-back confusion is often accompanied by a more diffuse perception of the sound source, and thus increases the uncertainty of estimating the position of the source. Both measures are illustrated in Fig. 5.1.

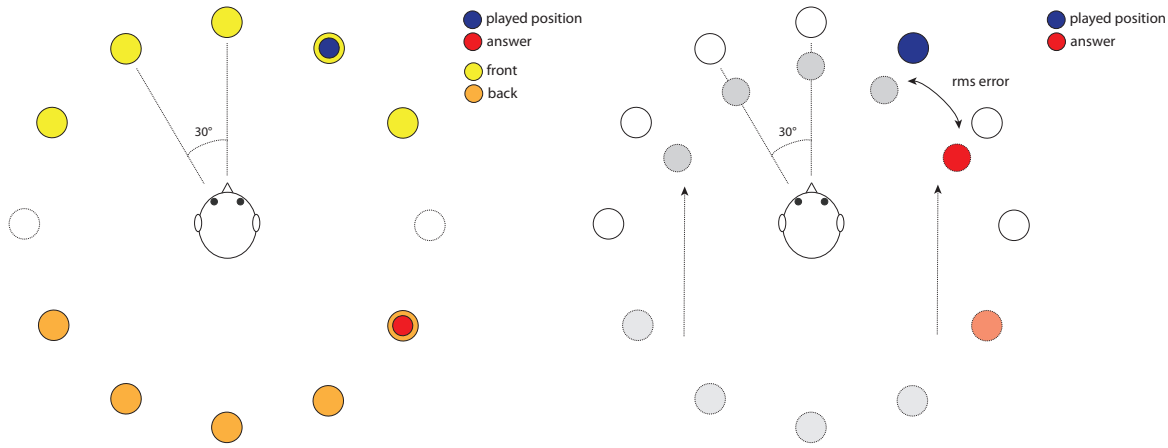


Figure 5.1: *Measures of localization performance: ♯ of front-back confusions (left) and directional RMS errors.*

The front-back confusions are counted and divided by the number of possible ambiguities. The score of a test subject for a test run is expressed in [%]. For positions played at 90° the confusions are not defined and are thus ignored. For any played position, responses at 90Âř are not counted as front-back confusions. Performing at chance level results therefore in a score of 41.66%. For computing the directional RMS errors, the front-back errors are first resolved by projecting the back positions to the front prior to the error calculation. This removes artificially large RMS values caused by a front-back confusion. A similar procedure was used to analyze the outcome of localization experiments in Chapter 3 and in [Langendijk & Bronkhorst 2000].

5.2.2.2. Analysis of head trajectories

For conditions with head movements, the head trajectories were recorded with the motion-sensor fixed on the top of the head of the subjects. To analyze the differences in head movements between normal hearing subjects and CI users, the trajectories were defined by the duration of the trajectory [s], the total length of motion [°], the number of movements towards the wrong direction and the trajectory complexity. The measures were applied to the head trajectories after a 3° movement from the initial and ending positions so that the reaction and response reporting times did not influence the results. The length of motion is defined as the total angular movement of a test subject. The movements towards the wrong direction count how many times the subjects displaced his head to the opposite direction of the target source by more than 5°. The trajectory complexity was estimated by fitting a polynomial function of increasing order to the curve until the error estimate of fit dropped under a threshold [Brimjoin *et al.* 2010]. The threshold was defined as a prediction error of 10% of the maximal value of the trajectory. Statistical significance of the results of head trajectory analysis was tested with the nonparametric Mann-Whitney U test.

The three first measures are illustrated in Fig. 5.2. Fig. 5.2 shows a head trajectory measured on one of the test subjects for a sound played at −60°. The green line shows the position played; the full trajectory recorded by the motion tracker is plotted in black. After an initial reaction time, the subject starts moving his head until he goes back to position and gives his response. The measures analyze the head trajectories after a 3° movement from the initial and ending positions (the blue curve) so that the reaction and response reporting times do not influence the results.

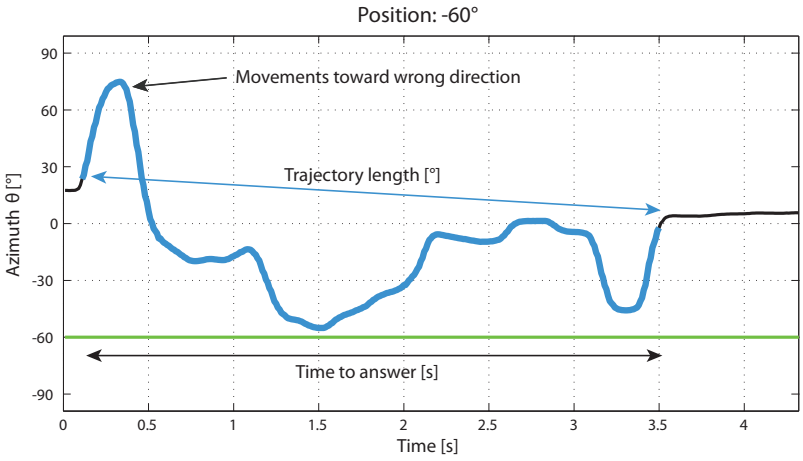


Figure 5.2: *Measures of head trajectories.*

5.3. Results

5.3.1. Analysis of localization performance

The outcome of the localization for the angular RMS errors experiment is shown in Fig. 5.3 Fig. 5.4 displays the number of front-back confusions. As expected, the CI users performed worse than the normal hearing subjects for all test conditions for both performance measures. It appears however that head movements did not improve the angular acuity (RMS errors) of the CI users (Fig. 5.3). The mean RMS errors were 30.6°, 26.9° and 27.6° for the short, middle and long signal lengths respectively. These differences are not significant. Normal hearing subjects show statistically significant lower RMS errors for the middle and long signal durations when head movements were allowed (0.26° against 2.05° and 0° against 1.7° for the middle and long durations respectively). For the angular RMS errors, significance was tested with Student's t-test.

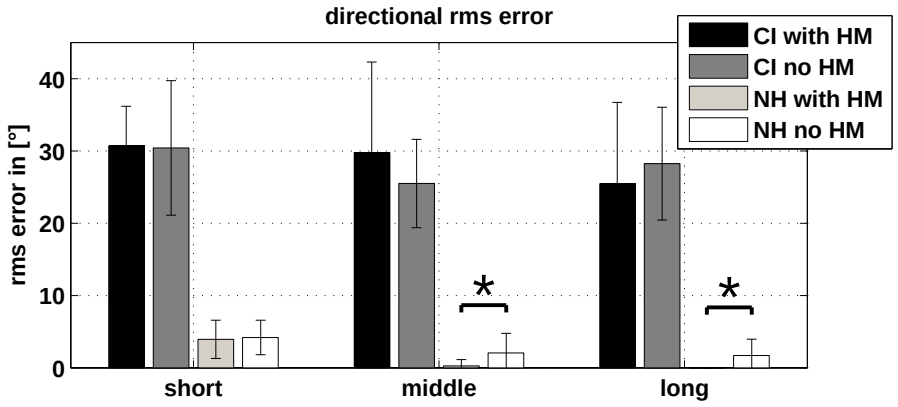


Figure 5.3: *Average RMS localization error for the normal-hearing and the CI subjects for the three signal lengths.*

For the CI subjects and the conditions without head movements, a similar percentage of front-back confusions occurred (around 25 %) independently of signal length (Fig. 5.4). This contrasts with the data from the normal hearing subjects, where confusions decreased with increasing signal duration. For the normal hearing listeners, the differences between the front-back confusions for the middle and long sentences were significant ($p = 0.007$). Statistical significance was tested with the nonparametric Mann-Whitney-U test. For the short signals, a trend can be seen which is not significant. During the "fixed head" conditions, head movements were monitored. It appeared that, while the listeners were instructed to keep their head still, small head movements between three to five degrees occurred. Those movements might have helped the normal hearing subjects but not the bilateral CI users and could explain the reduction of front-back confusions.

For all listeners, head movements significantly helped to distinguish between front and back positions provided the target sentences were long enough. For the CI users, the amount of

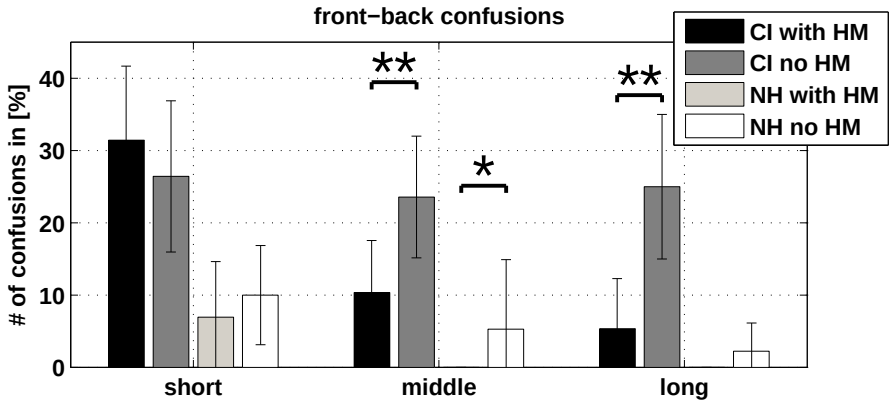


Figure 5.4: *Percentage of front-back confusions for normal-hearing and bilateral CI subjects.*

front-back confusions dropped from 23.6% to 10.4% and from 25% to 5.5% for the middle and long signals respectively. These differences are highly significant ($p < 0.01$). While listening to the short signals, the listeners were encouraged to move their heads even if it appeared counter-productive. Indeed, some CI users reported that head movements disturbed more than helped, because of the short duration of the test signal. They indicated more front-back confusions with head movements than without (31.4% against 26.4%). This effect is however not significant ($p = 0.22$). For the normal hearing subjects, head movements removed all front-back confusions for the middle and long sentences. For the short signals, there is a trend indicating that the normal hearing listeners could have used head movements for a better performance. However, the difference is not significant ($p = 0.06$).

5.3.2. Analysis of head trajectories

For conditions with head movements, Fig. 5.5 shows the trajectories of a normal hearing subject for all the positions tested. For this figure, test and retest sessions were combined. Every position was therefore played four times in total. In this particular example, the long signals were used.

The trajectories of normal hearing subjects are relatively similar for sounds played from the same location. After an initial reaction time, they moved their heads towards the position of the source without hesitation. This differs significantly from the head trajectories of bilateral CI users. Fig. 5.6 shows the head trajectories of a CI test subject listening to the long target signals for all positions. It appears immediately that the trajectories are much larger than for the normal hearing listeners. In this case, the CI test subject moved his head towards the left and the right for every position played. While the normal listeners could easily spot the right position, the CI users had to search the acoustical space to get an appropriate impression about the source position. Even in this condition, where the target signal duration is above four seconds, the movements and the responses were far off the sound source position.

For all measures, the CI subjects performed worse than the normal hearing listeners. The

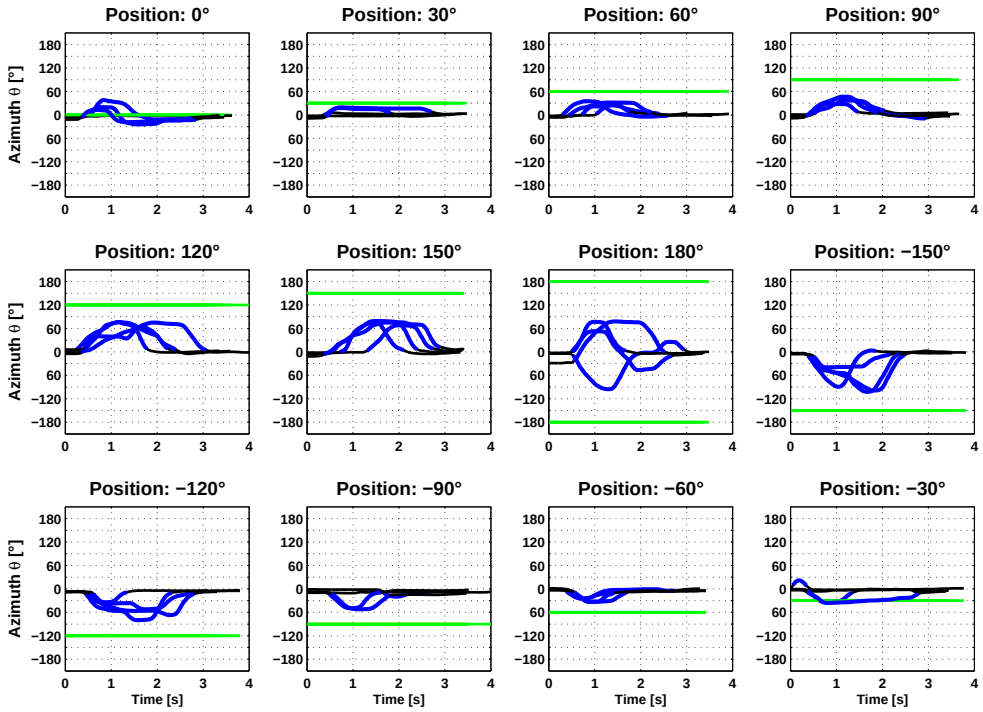


Figure 5.5: *Example of trajectories of a normal hearing subject for the long signal durations. The location of the played position is shown in green.*

total length of movement (Fig. 5.7) was greater for CI subjects for the middle and long target signal durations than for the normal hearing. This was significant for the long signals only ($p = 0.10$ and $p = 0.002$ respectively). No differences were found for the short signals. For the normal hearing subjects, the trajectory range and length did not differ between middle and long stimuli. This indicates that increasing the duration of the stimulus did not provide more useful information. A similar pattern can be seen when looking at the response delay of the test subjects (Fig. 5.8). The CI users score worse for all conditions. Statistical significance was reached only for the middle and long signals ($p < 0.001$).

When considering the movements towards the wrong directions (Fig. 5.9) the hesitation of the CI users is clearly visible. They rotated their heads in the wrong direction for all test conditions (Fig. 5.10). This effect is stronger when listening to long signals This behavior was rarely observed for normal hearing subjects. Some listeners reported focusing on the target loudspeaker with the appropriate ear and used the difference in signal to noise ratio as a cue to localization.

The polynomial order of the head trajectories was higher for the bilateral CI users (Fig. 5.10 for all signal lengths). This confirms the findings in Fig. 5.7. The difference between the trajectory complexities is statistically significant for the middle and long durations only.

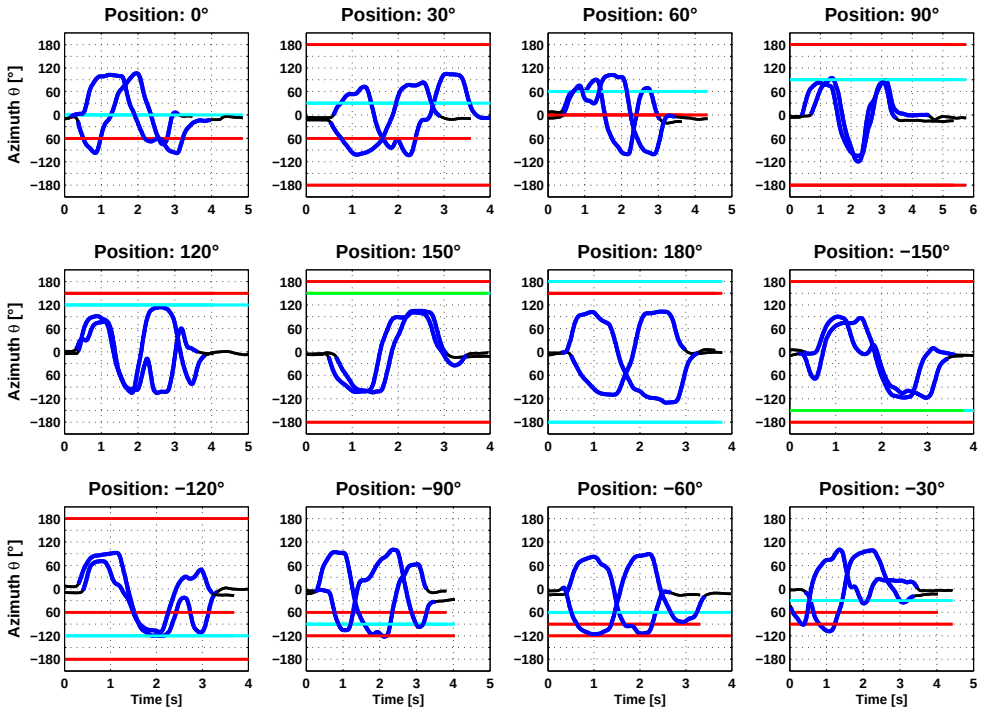


Figure 5.6: *Example of trajectories of bilateral CI users. The played position is indicated in cyan, a correct answer in green and wrong responses in red.*

The short signals were not long enough to measure a difference in this measure ($p = 0.09$). The test-retest analysis did not show any training effect between both sessions in any of the performance measures.

In Chapter 2, Fig. 2.5 we showed an example of the ILDs and ITDs of a human subject measured in an anechoic room. When looking at the figure, we see that for this specific subject, the broadband ILDs increase linearly for angles between 0° and 100° and reach a maximum value of 26 dB at 110°. The mean slope between 0° and 100° is $0.26 dB/°$. The average movement velocity for the CI users for the short stimulus duration was $35°/sec$. This implies that for a signal length of 503 ms, the maximal possible variation in ILD was around 4.6 dB, which is well in the range of detectable change in ILDs [van Hoesel 2004]. The same calculations for ITDs indicate that the CI users needed to detect a change of 0.135 ms. This value is in the range of the ITDs jnds of the best bilateral CI users van Hoesel found in [van Hoesel 2004].

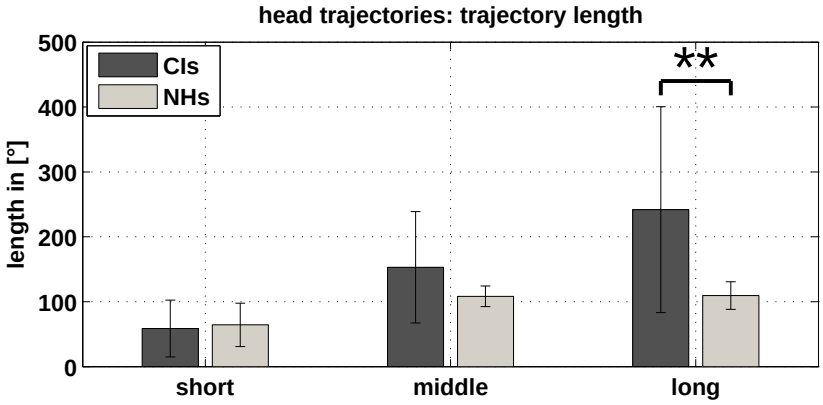


Figure 5.7: *Length of head trajectory in [°] for the CI users (dark grey) and the normal hearing subjects for the three signal durations.*

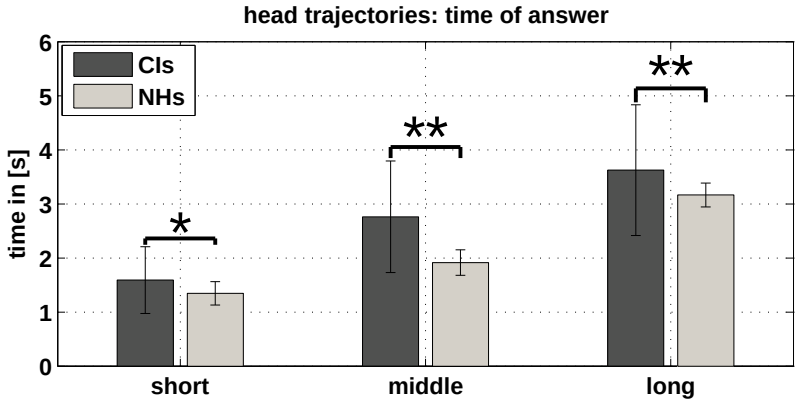


Figure 5.8: *Duration of head trajectory in [s] for the CI users (dark grey) and the normal hearing subjects for the three signal durations.*

5.4. Discussion

The results presented in this chapter confirm the hypothesis that head movements are essential for bilateral CI users for distinguishing between sounds played from the front and the back. In the situation with the longest target signals, the proportion of front-back confusions was reduced from 25% to 5.5%. This score is of the same order than the performance of normal hearing subjects for signals of medium length with fixed head position. This suggests however, that some confusions could not be resolved, even with very long listening time. For the long signals, the duration was 4.45 s. in average, which is long enough to scan the entire loudspeaker ring with head movements and focus on the appropriate loudspeaker.

Perrett and Noble [S.Perrett & Noble 1997] investigated human sound localization with and without head movements. The stimuli were white Gaussian noise of 0.5 and 3 seconds.

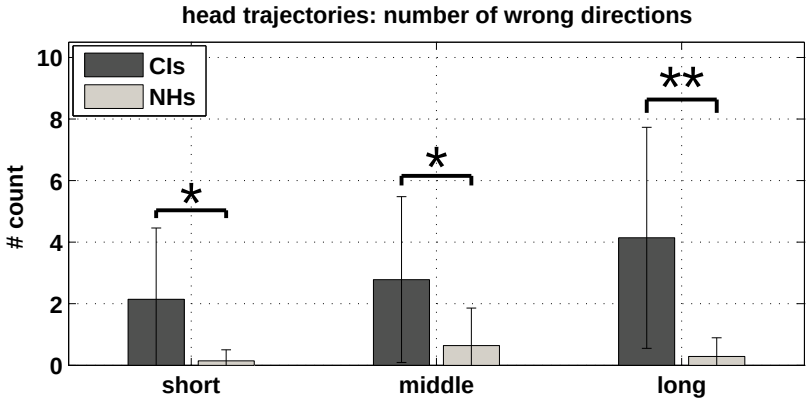


Figure 5.9: *Figure 6: Number of head movements towards the wrong direction (left-right) for the three signal durations.*

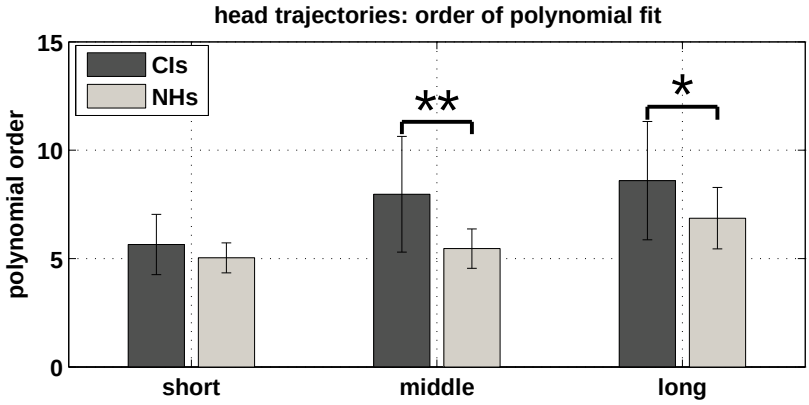


Figure 5.10: *Complexity of head movements measured as the order of the best polynomial fit to the head trajectories.*

They tested three conditions: motionless, natural movements and a forced movement towards 28Âř at stimulus onset. For the short signals, the proportion of front-back errors went from around 25% to 16%. By measuring the head trajectories, they observed that when the movements were larger than five degrees, no confusions occurred. This indicates that the normal hearing listeners can use head movements in signals as short as 500 ms, which is the duration of the shortest stimuli in our experiment. In our experiment with the short stimuli, the head movements did not significantly improve performance for the normal hearing subjects. Although a trend was seen for the rate of front-back confusions, it was not significant probably due to the limited number of test subjects or the presence of noise. For the 3 s. signal a few confusions occurred but only in situations where the listeners did not rotate their heads. In our study, the test subjects were encouraged to move their heads and could resolve all front-back ambiguities.

The head movements and the signal duration had practically no effect on the angular RMS errors of the CI users. This is quite surprising, as we expected for the long signals at least, an increase in performance as the listeners had enough time to screen the room and search for the target sound source. The error was around 28° for all conditions and was significantly higher than what other studies reported. Van Hoesel and Tyler [van Hoesel & Tyler 2003] for example found an average RMS error of 9.8° in quiet. The presence of background noise could explain this difference, although the target signal was always clearly audible and reported as such by the test subjects. Seeber and Fastl [Seeber & Fastl 2008] have suggested that the CI users primarily use differences in level for localizing sound sources. Furthermore, just-noticeable-difference (jnd) studies show ITD jnds of around 1 ms, which is above the ITDs useful for localization. The background noise could have masked the speech signal in the contralateral ear in regions where the head shadow effect was large and thus reduce performance.

In the study of Seeber and Fastl [Seeber & Fastl 2008] two bilateral CI users had to localize low-pass noise with a cut-off frequency of 500 Hz. One of the subjects tested performed well (RMS error of 8.1°). Based on Head-Related Transfer Functions (HRTFs) measurements, they observed ILDs up to 5 dB in this frequency region. To rule out that localization was based on low frequency ITDs, they set up a second experiment in which the localization cues available were either amplified ITDs or ILDs. In the ITDs only condition, the same test subject showed poor localization abilities. They concluded that the position of the low-pass noise of their first experiment was primarily estimated using ILDs only. Here, the CI users could not take advantage of the ILD fluctuations induced by their head movements as the RMS errors did not decrease with increased signal duration.

By comparing head movements of normal hearing and hearing impaired listeners in localization tasks with and without visual input, Brimjoin et al. [Brimjoin *et al.* 2010] observed that the most significant difference between the orienting responses of both subject groups was the complexity of the head trajectories. Among other factors, it appeared in their study that the movements of the hearing impaired were characterized by rapidly changing velocities, direction reversals and corrections of fixation positions. According to their conclusions, this can be explained by either the increased uncertainty in sound localization for hearing impaired or by a compensatory strategy to extract the most information possible from a given situation. In our study, the analysis of the head trajectories of the bilateral CI users showed similar characteristics.

5.5. Conclusion

Head movements contributed to sound localization of bilateral CI users, provided the stimulus was long enough. The main benefit was a reduction of front-back confusions. The angular acuity was however not improved. Normal hearing subjects on the other hand showed better performance in both measures.

The uncertainty in sound localization of the bilaterally implanted subjects was clearly visible in the more erratic characteristics of their head trajectories. In all trajectory measures,

they scored worse than the normal hearing subjects. The total length of movement and the order of polynomial fit clearly described the hesitating behavior of the CI users.

Head movements can have a positive effect on speech understanding as well, as mentioned in the introduction. It is however still unknown how large this benefit is for bilateral cochlear users, especially in more complex acoustical settings, such as noisy and multi-talker environments. The actual benefit which CI users extract from their devices would probably be higher than shown by standard clinical speech intelligibility tests. Further experiments that combine head movements and visual information in various acoustically challenging environments might be helpful to elucidate the real-life benefit of bilateral CIs.

6. Distance perception with bilateral hearing aids

6.1. Introduction

Auditory distance perception in humans is based on a variety of cues. While it appears that it is mainly based on vision, the auditory system can provide essential information in situations where the visual cues are incomplete or missing. The primary acoustic cue is the intensity of the target stimulus. In anechoic conditions and for point sources, the intensity of sound follows the inverse-square law and decreases by 6 dB for every doubling of distance. In reverberant environments, the law does not hold but sound intensity decreases with increasing distance nonetheless (see for example Fig. 6.2).

The ratio between direct and reverberant energy is another cue for the perception of distance of the source in reverberant environments. When the source is close to the listener, the direct energy dominates. As the source moves away from the listener, the diffuse components of the room impulse response begin to dominate (Fig. 6.2). The Direct-to-Reverberant Ratio (DRR) depends on the acoustical qualities of the environment. Contrary to the intensity cue, the DRR allows an *absolute* judgment on distance. It has been shown that the importance of the DRR in distance perception is relatively small compared to the intensity cues . Nevertheless, a study by Mershon and King [Mershon & King 1975] has shown that the perceived distance of a sound source is more accurate in reverberant than in anechoic environments. This suggests that although secondary the DRR is an essential cue to sound distance perception.

For large distances, changes in the spectrum of the sound source can also affect distance perception. The sound absorption properties of air modify the spectrum of the sound at the ears. Specifically, high frequencies are more attenuated. This effect is however relatively small. It is in the order of 3 to 4 dB per 100 meters for a frequency of 4 kHz [Ingard 1953]. Similarly to the absolute intensity cue, the spectral cue needs an a priori estimate of the source spectrum in order to make a decision. The importance of this cue is relatively small compared to the intensity of sound and the DRR. Studies have nonetheless shown that spectral cues play a role in distance perception.

Interaural time and level differences were extensively discussed in previous chapters of this work. It was shown that they provide essential information for sound localization, source width and internalization. It appears that the binaural cues also carry information about the absolute distance of a sound source. For distances smaller than 1 meter, measurements of interaural level and time differences revealed strong variations of interaural level differences

for approaching sources whereas the interaural time differences are only slightly modified. This could be used by the human auditory system to detect small variations in distance for close sources. Brungart et al [Brungart *et al.* 1999] examined the perceived distance of nearby sources. DRR and intensity cues were made inoperant by carrying out the experiment in anechoic conditions and by roving the intensity of the stimulus. The test subjects were blindfolded, to avoid the influence of visual cues. In these conditions, the listeners were able to make relatively accurate judgments about the distance of the stimulus when it was played at 90^o. However, when the sound came from 0^0, their performance dropped significantly. Repeating the experiment with a low-pass and a high-pass filtered white noise, it appeared that the low-pass stimulus could not be localized in contrast to the high-pass stimulus. These results suggest that interaural level differences are distance cues for nearby sources. For sources further away from the listener, these effects are negligible.

Virtual acoustics have been used to test human distance perception as well. In a recent study, Zahorik [Zahorik 2002] used individually measured Binaural Room Impulse Responses (BRIRs) for positions in the front and at the right of the listeners (0^o and 90^o) in a reverberant space (an auditorium). Twelve locations were measured (0.3-13.79m). Noise and speech signals were filtered with the individual BRIRs and presented to the listeners over headphones. The results show that listeners consistently underestimate the absolute distance of the virtual sounds. The study further examined the listeners weighting of the intensity and DRR cues for distance perception. The advantage of virtual acoustics is that the BRIRs can be manipulated so that one of the two cues can be dominant. This was done by scaling the BRIR for intensity and modifying the level of the direct-component only for the DRR. What the results show is that the cue-weighting strategy followed by the listeners was different for both stimuli but was not dependent on position. For the speech signal, the dominant cue was the intensity of the stimulus.

Distance perception with hearing aids has been addressed in the Speech, Spatial, and Qualities of Hearing (SSQ) questionnaire. Seven questions dealt with distance perception in real-world environments. The questions asked about the expected distance of a sound ("*Do the sounds of people or things you hear, but cannot see at first, sound closer than expected?*"), the distance of moving objects ("*Can you tell from the sound whether a bus or a truck is coming towards you or going away?*") and the expected location of a sound ("*Do you have the impression of sounds being exactly where you would expect them to be?*"). The questionnaire was used by Noble and Gatehouse ([Gatehouse & Noble 2004, Noble & Gatehouse 2006]) to evaluate the abilities of hearing impaired subjects prior to fitting, unilaterally fitted and using bilateral hearing aids [Noble & Gatehouse 2006]. The answers to the distance-related questions suggest that there is a significant improvement with bilateral hearing aids compared to unilateral and no fittings conditions. The amplification provided by one hearing aid however seems not to be sufficient to improve distance perception. The question arose if compression algorithms could distort distance perception of hearing aid users because they modify the levels of incoming sounds. In a recent study [Akeroyd 2010], no significant difference in Just-Noticeable-Differences (JNDs) was observed for different compression ratios. The results could be explained by the fact that the hearing aid users were trained to the effects of compression on sound intensity for sound perception.

In this project, the influence of four hearing aid algorithms on distance perception was investigated. The same algorithms are used as in the localization experiment discussed earlier (see chapter 4). Distance perception was tested because the algorithms could potentially distort distance cues: the omnidirectional and directional microphones act on the reverberant part of the signal by picking up more (omni) or less (beam) reflective energy from the back. This modifies the DRR. The noise canceler modifies levels on the left and right hearing aid independently which might influence the intensity cue. The DRR might be modified as well by the noise canceler, as the reverberation noise is reduced by the algorithm.

6.2. Methods

The experimental framework of Akeroyd et al [Akeroyd 2010, Akeroyd *et al.* 2007] was slightly modified for the purposes of this experiment. The system for virtual acoustics described in Chapter 3 was used to reproduce a virtual room in which the listeners had to guess the apparent distance of sentences spoked by a male or female speaker. The simulated room was the same as in the experiments of Akeroyd et al. It was rectangular, 7 meters wide, 9 meters long and 2.5 meters high. The absorption coefficients of the walls were set to 0.5. The floor and the ceiling didn't reflect sound. The walls were perfectly reflexive, which implies that all the reflections were specular, with a reduction of 3 dB per surface hit. Using Sabines' law, the reverberation time was estimated at 250 ms. The BRIRs were simulated using the ROOMSIM software [Campbell *et al.* 2005] using individually measured HRTFs. Contrary to Akeroyd et al's experiment, the sounds were reproduced using the open CIC speakers and not through a ring of loudspeakers.

The stimuli were played in pairs. One sentence was spoken by a male, the other by a female. The task of the test subjects was to find out which one of the two stimuli was the furthest away. One of the signals was always at a reference distance. Two reference distances were chosen: 2 and 5 meters. The comparison distances were distributed on a line, 30^{o} from the listeners and were either closer or further away from the reference. In the setup described by Akeroyd et al, the sources were always played in the front of the listeners. It was decided to present them slightly on the side to reduce the risks of sound internalization. The setup of the experiment is shown in Fig. 6.1. The reference distances are shown in yellow.

For the 2-meters reference condition, the comparison distances were set to 1, 1.33, 1.66, 2.33, 2.66 and 3 meters. For the 5-meters reference condition, the comparison distances were set to 2, 3, 4, 6, 7, and 8 meters. One of the two stimuli was set randomly at the reference distance. The order of the signal pairs was set randomly. To prevent the listeners from detecting small changes in the stimuli and to learn which of the signals was played at the reference positions, one of the presented stimuli was spoken by a female and the other by a male. The sentences were taken from the Basel sentence test material [Tschopp & Ingold 1992]. The signals were played at 60 dB at the reference positions. They were calibrated based on their rms values. To challenge the algorithms and to simulate more realistic conditions, a diffuse background noise was played as well. The SNR was set to 5 dB.

The direct-sound and the reverberation components of the BRIRs can be directly com-

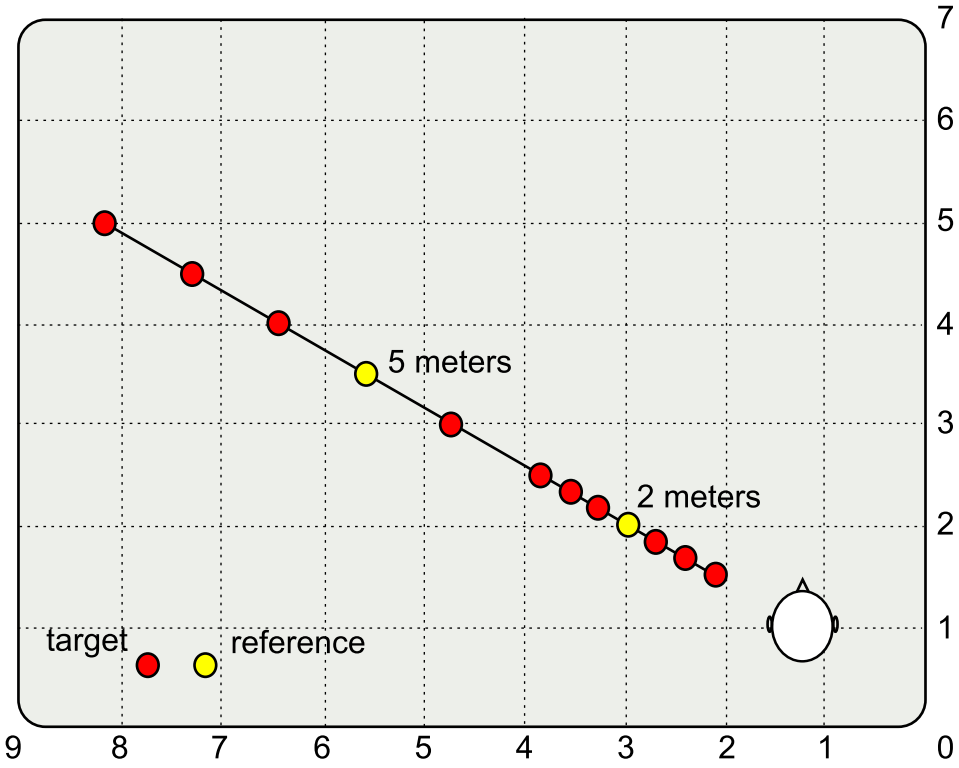


Figure 6.1: *Experimental setup: virtual playback room with the positions simulated on a line* 30^o *on the left of the listener.*

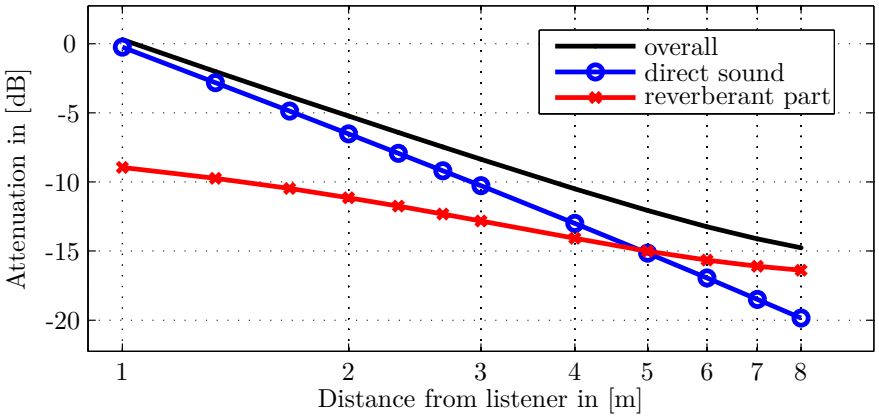


Figure 6.2: *Level decay of the BRIRs (black), direct sound (blue) and reverberant part of the impulse response (red).*

puted from the simulated impulse responses. They are displayed in Fig. 6.2. For positions close to the listeners, the energy decay of the BRIR follows approximately the inverse square

law. The further the sound moves away from the listener, the stronger is the reflective sound field and the weaker is the intensity cue. The evolution of the DRR with respect to distance can also clearly be seen in this picture. For distances above 5 meters, the reflective energy dominates the BRIR.

Twelve normal hearing listener took part in the experiment. They had their hearing measured by standard audiometry prior to the experiment. No hearing loss above 20 dB HL across all frequencies was detected.

6.3. Results

The outcome of the distance experiment is shown in Fig. 6.3. The results are separated into four different conditions. For both reference distances (2-meter and 5-meter), the results are analyzed separately depending on whether the test stimuli were set at distances closer (2-closer and 5-closer) or further away (2-further and 5-further) from the listeners. The bars in Fig 6.3 represent averaged results across the 3 distance intervals. The mean percentage of correct answers and one standard deviation are listed in Table 6.1.

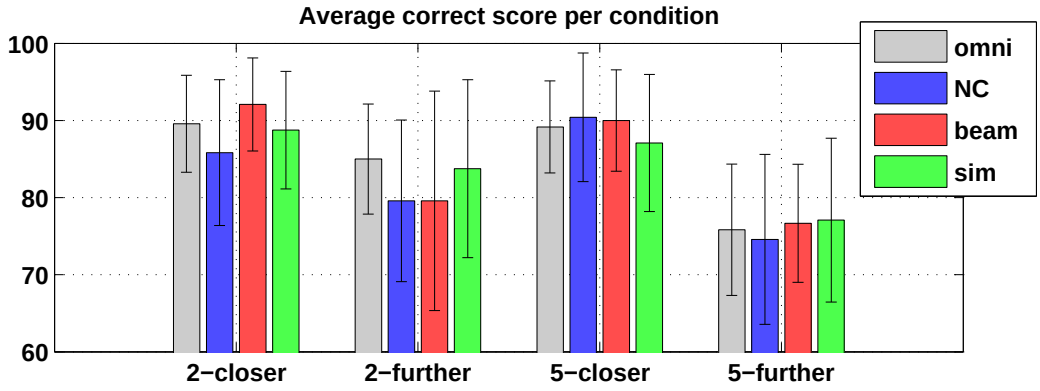


Figure 6.3: *Outcome of the distance test. The results are divided into positions played closer or further of the reference position for the 2 meters and 5 meters reference distances respectively. The average results for the four algorithms are shown.*

The results show that the algorithms did not have a significant effect on distance perception. The standard deviation is however very large for every tested condition (12.5 % in average). Increasing the number of test subjects could reduce this number. The variance of the results is not distributed equally across conditions. When the comparison distance was closer to the listener (2-meters closer and 5-meters closer) the standard deviation dropped to 9.7 %. This compares to 15.4 % for the two conditions where the comparison distance was played further away than the reference. This difference can be explained by the increased difficulty of the 2 and 5 meters further condition. The SNR was lower in these two settings, making the detection of the comparison more difficult. In the 5 meters further situation, the performance of the test subjects was rather low with 76 % of correct responses on average.

Table 6.1: *Average percentage of correct responses for the four hearing aid algorithms. The table displays the results for the two reference distances.*

	2 meters closer				2 meters further				5 meters closer				5 meters further			
correct [%]	omni	NC	beam	sim	omni	NC	beam	sim	omni	NC	beam	sim	omni	NC	beam	sim
mean	89.6	85.8	92.1	88.7	85.0	79.6	79.6	83.8	89.2	90.4	90	87.1	75.8	74.6	76.7	77.1
std	8.9	12.7	8.3	10.3	12.3	16.6	19.1	14.1	8.8	10.1	7.4	11.1	13.3	16.3	13.7	17.8

The results for every distance interval are shown below in Fig. 6.4 for the omnidirectional algorithm. The figure displays the percentage of correct responses for absolute distance differences (left panels) and relative distance changes (right panels). The results for the other three algorithms follow the same pattern and are therefore not shown here. As expected, the smaller the distance difference, the harder it is to tell which of the two signals is closer. Fig. 6.4 confirms the finding that the 5-further condition is the most difficult condition.

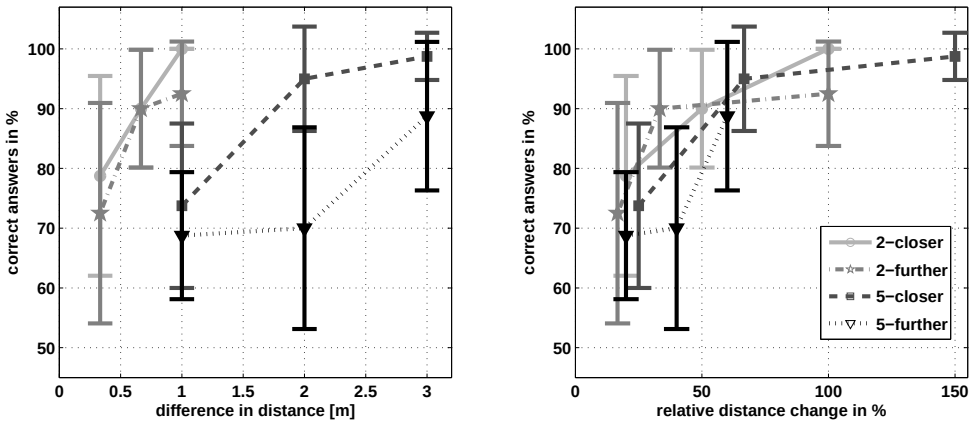


Figure 6.4: *Average number of correct scores for the omnidirectional algorithm. On the left, the results are shown in relation with the exact distance differences. On the right, the distance judgments are shown relative to the reference distances.*

Interestingly, when plotted with relative distance changes, there are few differences between the results for the 2 and 5 meters reference distance conditions. Still, in this condition the differences between 2- and 5- further conditions appear to be large, although non significant.

To compare performance between algorithms, a psychometric curve that relates the percentage of correct responses to a relative change in distance can be computed. The curve is obtained for each algorithm by interpolating and averaging the tested conditions across all relative distances. The psychometric functions are shown in Fig. 6.5. The figure confirms the previous analysis. The performance of the four algorithms is similar. There is a trend however that indicates that performance with the noise canceler was worse for small distance changes.

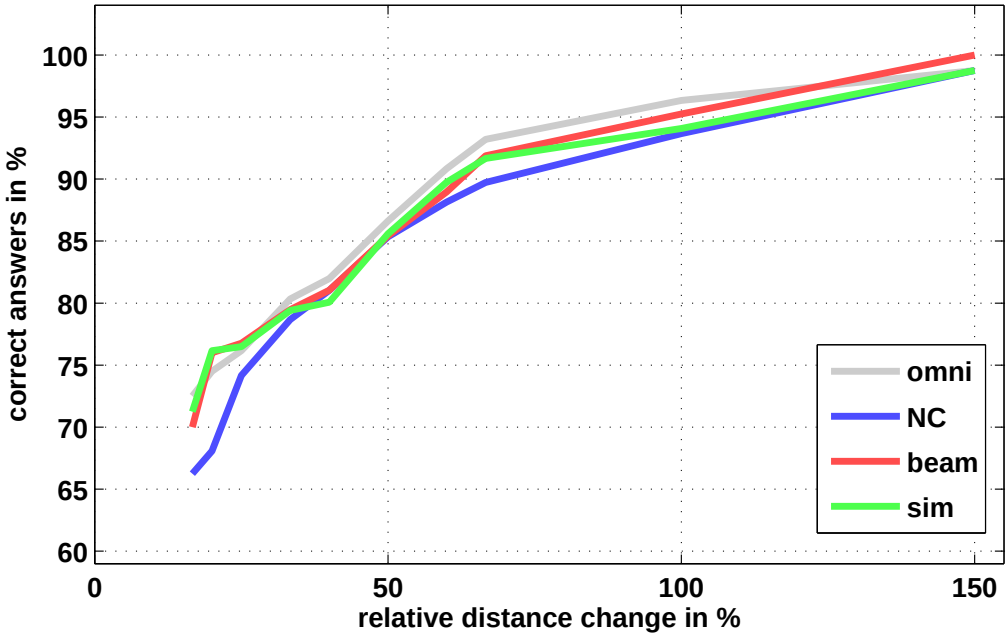


Figure 6.5: *Psychometric curves for the three algorithms evaluated and the reference (sim).*

6.4. Discussion and Conclusion

The results presented in this chapter are somewhat inconclusive. While there are strong differences in the spatial auditory representations of the sounds processed by the hearing aid algorithms, the data show that changes in sound distance are perceived with similar precision. The large standard deviations in the data might hide differences in performance. This high variation in the results is due to the small number of test subjects and the design of the test itself. The presence of background noise and the alternation of male and female speakers made the test more difficult and the responses of the listeners more uncertain. This effect could be reduced by testing with other stimuli and SNRs. However, the background noise has to be strong enough so that the algorithms work correctly.

The results confirm the findings of the SSQ questionnaire stating that the perception of sound distance with bilateral hearing aids is not an issue. The distance steps tested here were rather small (0.33 and 1 meters for the 2 and 5 meters reference conditions). For the smallest distance intervals, the listeners were able to identify the reference position in 70% of the cases. This implies that bilateral hearing aids offer an accurate reproduction of sound distance.

Distance perception with hearing aids was rarely experimentally investigated before. To the knowledge of the author, the study of Akeroyd [Akeroyd 2010], was the sole attempt aimed at measuring objectively how some aspects of hearing aid processing could affect auditory distance perception. In Akeroyd's study it has been found that hearing aid compression has

no significant effect on just-noticeable changes in sound distance even though the intensity cues are severely modified by the signal processing.

Part III.

Predicting and improving algorithm performance

7. Predicting spatial perception

7.1. Binaural Auditory System Simulator

7.1.1. BASSIM implementation

The Binaural Auditory System Simulator (BASSIM) has been designed to predict the perception of spatial features of the auditory space. It is based on a modified implementation of Breebaart's model for binaural detection. A schematic view of the implementation is shown in Fig. 7.1. It is composed of a peripheral part that models the processing of the outer and middle ear and the frequency decomposition of the inner ear. After being decomposed in critical bands, the signals reach the binaural processor where the interaural time and level differences are extracted. The binaural processor is composed of EI elements as shown in Fig. 2.14. The implementation of the peripheral and binaural stages of the BASSIM follow closely the description of Breebaart et al. [Breebaart *et al.* 2001] and are discussed in the following sections.

7.1.1.1. Peripheral model

The peripheral part of the BASSIM implements the processing done at earlier stage in the human auditory system. The first stage of the peripherical model simulates the processing of the outer and middle ears. It is implemented as a bandpass filter with a -6dB/oct roll-of below 1 kHz and 6dB/oct above 4 kHz.

The basilar membrane performs a frequency decomposition of the left and right signals as described in Chapter 2. It is performed using a third order gammatone filterbank as shown earlier. The bandwidth of the filters is equivalent to one critical band.

At the next stage of the peripheral processor, internal noise is included in the model in order to simulate the absolute threshold of hearing. The noise is implemented as white gaussian noise and is independent in the different channels.

The inner hair cells are modeled by applying half-wave rectification and low pass filtering to the input signals. The lowpass filter has a cutoff frequency of 770 Hz. That implies that for frequencies under 770 Hz the fine structure of the waveform is preserved. For frequencies up to 2000 Hz, the fine timing information of the signal is gradually reduced until the envelope only is transmitted. This procedure effectively simulates the gradual loss of phase information with increasing frequency that has been observed in the auditory nerve by various studies.

The original model of Breebaart includes an adaptation stage composed of a chain of five adaptation loops. They were included in Breebaart's model because they could explain binau-

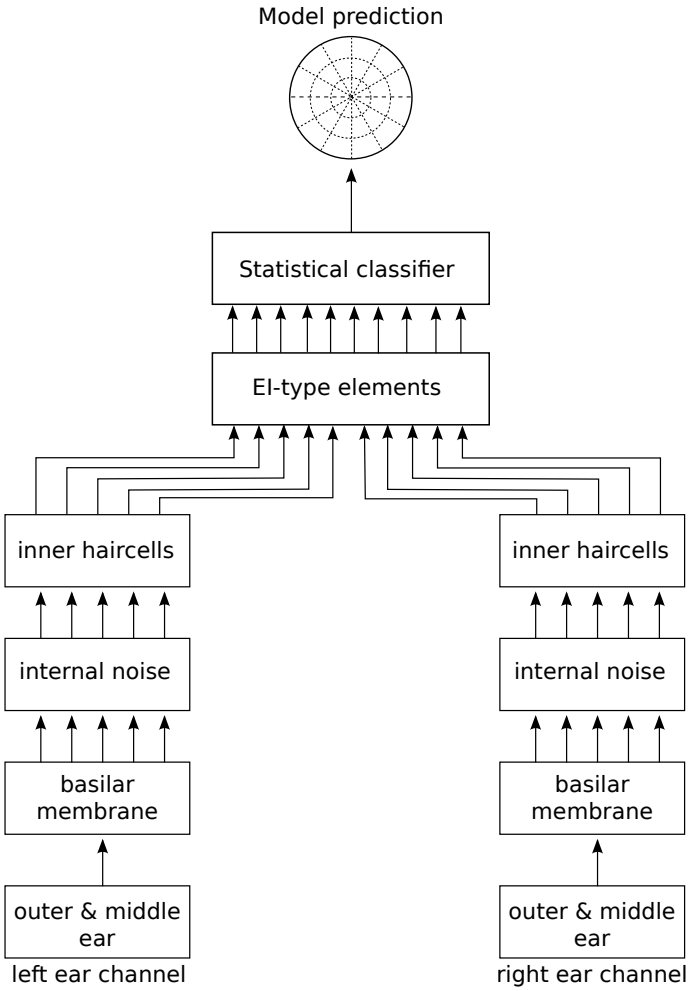


Figure 7.1: *Schematic representation of the binaural auditory system simulator*

ral forward masking data. The outcome of the adaptation loops is a logarithmic compression. This however changes the ILD of a constant binaural input and makes the localization of sound source based on the outcome of the model more difficult. This is the reason why we did not include the adaption stage in BASSIM.

The output of the peripheral model for two sine tones is shown in Fig. 7.2. In this figure, the peripheral model has been stimulated with tones of different frequencies set at 70 dB SPL. The output of a signal with a center frequency at 500 Hz is shown in blue. Simultaneously, the model was stimulated with a 4000 Hz tone. The output of the 4000 Hz channel is shown in red. The effect of half-wave rectification and low-pass filtering is clearly visible. For the 500 Hz tone, the fine structure of the signal is preserved. This contrasts with the 4000 Hz signal where only the envelope is transmitted to the higher stages of the model. The differences in level between the signals is due to the outer and middle ear model that attenuates more the 4000 hz signal. Here, the internal noise has been disabled.

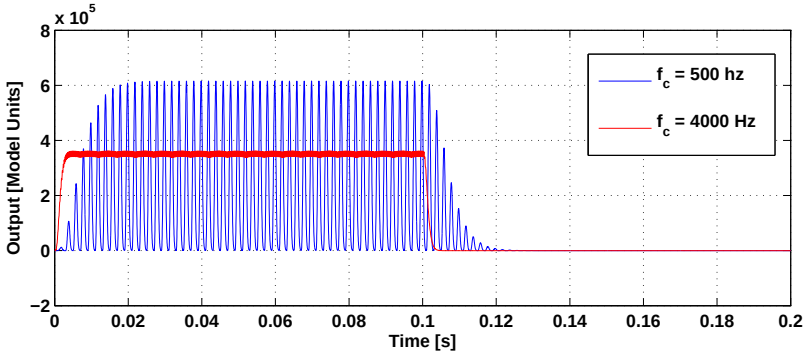


Figure 7.2: *Output of the peripheral model for two sine tones with center frequencies at 500 Hz (blue) and 4000 Hz (red). Two channels have been simulated with their central frequencies tuned at the frequencies of the stimuli. The signals were scaled at 70 dB SPL.*

7.1.1.2. Binaural processor

The structure of the binaural processor is schematically shown in Fig. 2.14. It takes as input the output of the left and right peripheral models. For each characteristic frequency, the signals pass through a delay line and a series of attenuators. The processor can be seen as an orthogonal combination of the delay lines of Jeffress ([Jeffress 1948]) and the model of the lateral superior olive of [Reed & Blum 1990] composed of a serie of gains. For each combination of discrete delays τ and gains α the binaural processor computes the difference between the shifted and attenuated left and right signals. At the left side of the binaural processor, the EI-elements are excited by the left ear signal and inhibited by the right ear signal. According to the description and implementation of Breebaart ([Breebaart *et al.* 2001, Breebaart 2001]), the output of the EI-elements is given by:

$$E_l(i,t,\tau,\alpha) = \lceil 10^{\alpha/40} L_i(t+\tau/2) - 10^{-\alpha/40} R_i(t-\tau/2) \rceil^2 \tag{7.1}$$

where $L_i(t)$ and $R_i(t)$ are the left and right outputs of the peripheral model for frequency band i at time t.

At the right side, the EI-elements are inhibited by the left ear signal and excited by the right input. This yields:

$$E_r(i,t,\tau,\alpha) = \lceil 10^{\alpha/40} R_i(t+\tau/2) - 10^{-\alpha/40} L_i(t-\tau/2) \rceil^2 \tag{7.2}$$

In Eq. 7.1 and 7.2, the $\lceil \cdot \rceil$ operator denotes half-wave rectification. This implies that the difference between the inhibitory and excitatory component is set to zero when negative. Eq. 7.1 and 7.2 result in a discrete sampling of the $\tau - \alpha$ space that correspond to different ITDs and ILDs. For each combination of ITD and ILD, the EI-elements of the binaural processor produce an output that is minimal for the $\tau - \alpha$ combination that corresponds to the ITD and ILD of the signal. [Breebaart *et al.* 2001] show that Eq. 7.1 and 7.2 are equivalent to:

$$E(i,t\tau,\alpha) = \left(10^{/alpha/40} L_i\left(t+\tau/2\right) - 10^{-\alpha/40} R_i\left(t-\tau/2\right)\right) \tag{7.3}$$

The limited temporal resolution of the binaural system has been included in the model by adding an integration window $\omega(t)$ to the equation. It is defined as:

$$\omega(t) = \frac{exp(-\left|t\right|/c}{2c} \tag{7.4}$$

where c is the time constant of the window. It is set to 30 ms based on findings by [Holube *et al.* 1998]. With this, Eq. 7.3 changes to:

$$E'(i,t,\tau,\alpha) = int_{-\inf}^{\inf} E(i,(t+t_i nt),\tau,\alpha)\omega(t_i nt)dt_i nt \tag{7.5}$$

Additionally, the binaural processor implementation of Breebaart [Breebaart 2001] includes saturation effects of the EI-elements and a weighting function $p(\tau)$ that emphasize central delays (τ close to zero). The saturation effects are modeled by a compressive function. This implies that the final output of the EI-elements is:

$$E''(i,t,\tau,\alpha) = ap(\tau)log(E'(i,t,\tau,\alpha)+1) \tag{7.6}$$

The weighting function corresponds to the centrality function of [Stern & Colburn 1978, Stern *et al.* 1988] . It is based on the assumption that more cells in the human auditory system are tuned to low delays. It has been successful in predicting various lateralization experiments in correlation-based models (see weighted-image model for example described in Chapter 2). The centrality function is defined as:

$$p(\tau) = 10^{-|\tau|/5} \tag{7.7}$$

The values of the parameters of the compressor (a,b) are set to 0.1 and 0.0002 according to Breebaart's description. Breebaart added an internal gaussian noise to the output of the EI-elements in Eq. 7.6. For simplification, the noise has been removed from the current implementation.

Four examples of the output of the binaural processor are shown in Fig. 7.3. Here, the frequency channel of 500 Hz has been simulated. The input stimulus is a 500 Hz sine tone. The ITD and ILD of the signal was varied for the four sample plots. It can be seen that when an ITD or an ILD is applied, the energy of the binaural output is minimum at the $\tau-\alpha$ position that corresponds to the ITD or ILD of the input stimulus. For ITDs different from zero, the minimum of the binaural output repeats itself at $\tau = 1500\mu$s. This due to the periodicity of the cross-correlation function that is effectively caused by the delay lines. By restricting the maximum range of $[-\tau;\tau]$ this can be avoided. Additionally, the weight of the centrality function can be increased in order to emphasize small delays. The width of the local minimum is related to the interaural correlation. The smaller the interaural correlation, the wider the peak of the local minimum. In the original model of Breebaart, the output of the

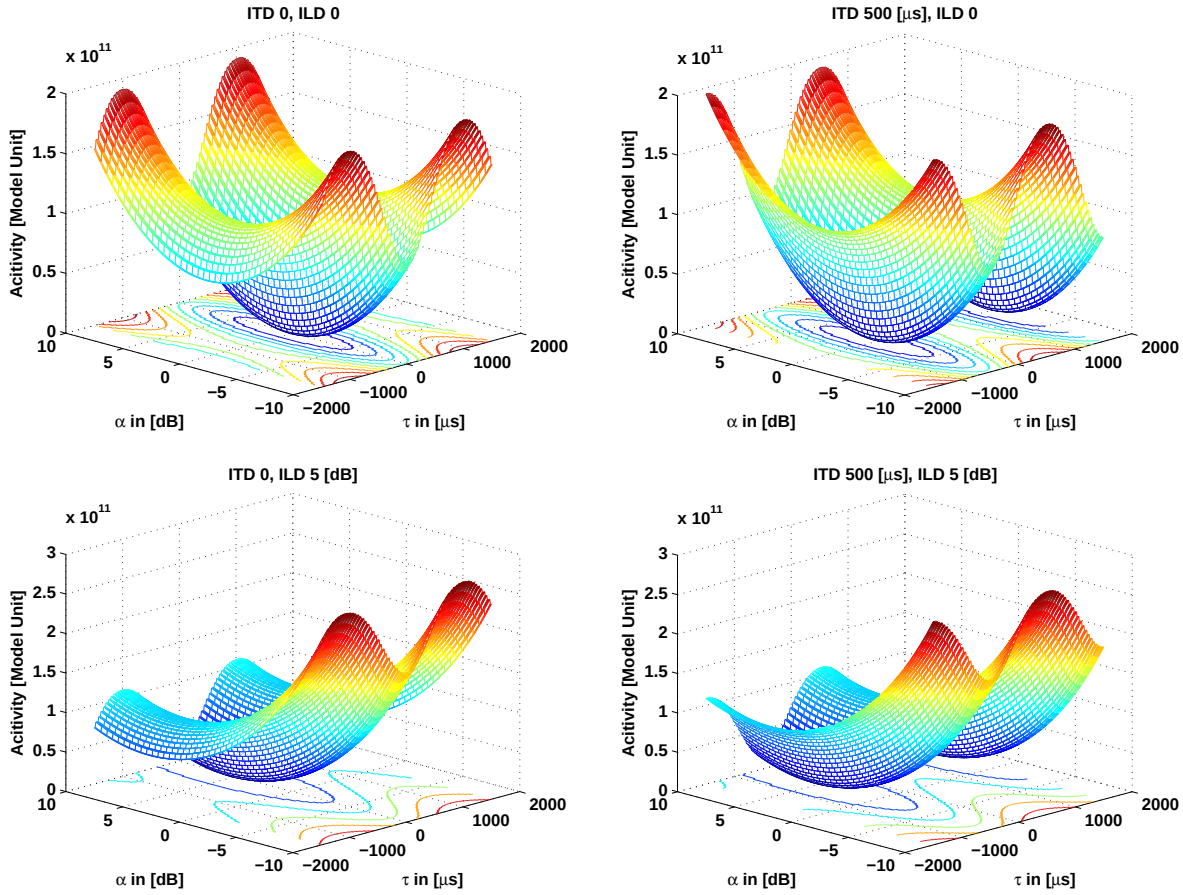


Figure 7.3: *Output of the binaural processor for a 500Hz sine tone. The ITD and the ILD has been set to O in the upper left figure. The ITD was set to 500 μs in the figure on the upper right corner. Below, on the left the ITD is set to 0 and the ILD to 5dB. Finally, down in the right the ITD is 500 μs and the ILD 5 dB. The detected ITD-ILD pair corresponds to the minimum of the function in the $\tau - \alpha$ space.*

whole binaural processor is fed to higher stages of the model. For reducing the computational power, only the τ and α values that correspond to the minimum of the binaural output are fed to the next stages.

7.1.1.3. Statistical classifier

The output of the binaural processor for a large number of frequencies is processed by a statistical classifier. The classifier is based on a random forest implementation [Breiman 2001]. It has been trained on measured HRTF data that are combined with KEMAR data as described in Chapter 3. In this section, a short introduction to random forest and the training procedure is given.

Decision trees Decision Trees is a classification method used to give a prediction on an output variable based on a set of input observations. Decision trees are binary, that is each node has either two leafs or none. They are separated into classification and regression trees. Classification trees are used to assign the input variables to a set of defined classes whereas regression trees assign the data to a set of scalar values.

At each node a test is applied to one of the input variables (x_j). The test, or splitting rule, corresponds to a yes/no question which reduce at each stage the dimension of the data. Depending on the outcome of the test, we move either on the left or on the right sub-branch of the tree. At the end, when reaching an end-node (leaf), a prediction can be made. The prediction averages all the training data points that reach this leaf node.

The decision trees building algorithm, called CART, was proposed by [Breiman 1984]. It consists of the following steps: Let us consider a vector of observation samples X composed of M variables x_j to which is assigned a response vector Y. At each node, the splitting tries to minimize the expected sum of variance of the following nodes. Mathematically, the splitting rule can be written as:

$$\text{argmin}_{x_j \leq x_j^R, j=1,...,M} \left[P_l \text{Var}\left(Y_l\right) + P_r \text{Var}\left(Y_r\right) \right] \tag{7.8}$$

with Y_l and Y_r the response vectors of the left and right child node respectively, x_j^R the best splitting value of variable x_j. $x_j \leq x_j^R, j = 1, ..., M$ is the optimal splitting question with P_l and P_r the probability of moving to the left of respectively right child node.

Random forests Random forests as classification method were proposed by Breiman [Breiman 2001]. They can be described as an ensemble of B trees $T_1\left(X\right), ..., T_B(X)$, where X is the vector of input variables as above. The trees produce B outputs $Y_1 = T_1(X), ..., Y_B = T_B(X)$ that correspond to the output of each tree in the random forest. The final prediction $\hat{Y}$ is the average of all the individual predictions. The training procedure is based on the following steps. First, from the training data n_{tree} bootstrap samples are taken (i.e the input data is randomly sampled with replacement). For each bootstrap sample, a regression tree is grown. At each tree chose the best split based on $m_{try} \leq M$ variables. n_{tree} and m_{try} are parameters of the random forest generation algorithm. The procedure is repeated until n_{tree} trees are grown.

Training of the classifier The training and the prediction of the regression classifier is done with the original MATLAB interface provided by Breiman and Cutler* [Breiman 2002]. A random forest was generated for all 710 positions that are covered by the KEMAR HRTF data set (see Chapter 3). Additionally, a model has been trained for a set of central frequencies $f_c = [125, 250, 500, 1000, 1500, 2000, 4000, 5000]$. According to the description of the binaural processor in section 7.1.1.2, the variables transmitted to the classifier are the τ and α of the minimum response of the EI-elements. m_{try}, or the number of variables used for choosing the best split, is set to 2. To generate the training data, a 200 ms white noise was filtered with all

*Available at http://www.stat.berkeley.edu/users/breiman/

HRTFs. For each position, the signals were then processed by the binaural model, yielding a set of τ and α. To generate more training data and to cover more conditions, the input of the binaural model was scaled at different sound pressure levels ($L = [20, 30, 40, ..., 110]$) before being processed by the model. For a model, half of the input training vector X was composed of the τ and α that resulted from noise filtered with the HRTF of the corresponding position. The other half was composed of τ and α resulting from the processing of random positions and levels. In total, the input training data was composed of 4000 samples. The value 1 was assigned to the target output vector Y for data corresponding to the model positions and 0 for the random inputs. The number of trees in the model was arbitrarily set to 500. Fig. 7.4 shows the regression tree grown to model sound coming from 90° in the horizontal plane.

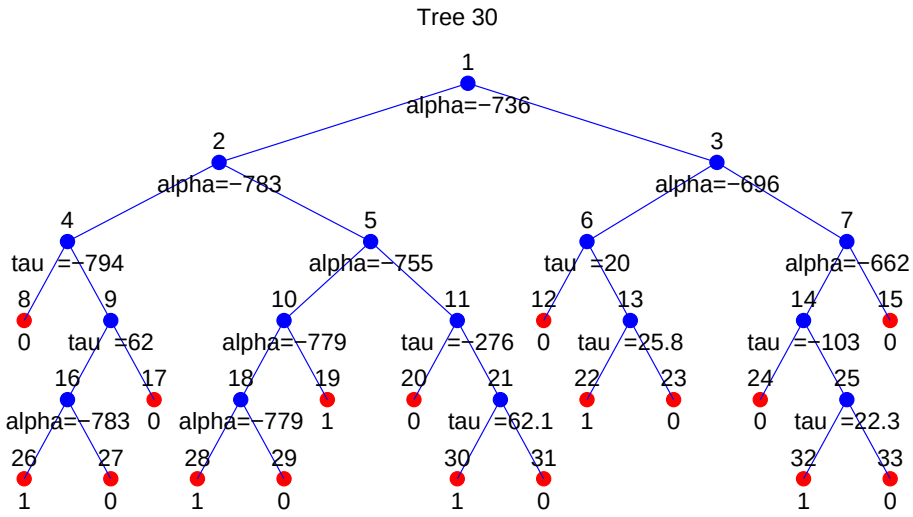


Figure 7.4: *Tree for a simulated position at* 90^o.

7.2. Prediction of spatial perception

Once the model had been generated for all the simulated positions, a prediction on new data can be made by running the signal through the binaural processor and through the classifier. The perception of the stimulus is then interpreted based on the energy distribution of the output over all models. The classifier is run on each frequency band separately and the results are averaged across frequency bands. Fig. 7.5 shows an ideal perceptual map generated for a white noise coming from 30° in the horizonal plane. The models have been trained using measured HRTFs on the same subject. Despite the conditions being ideal (anechoic, no interfering signals) the cone of confusion is visible (dark brown). Nevertheless, the BASSIM shows a clear compact source located at the correct position.

The accuracy of BASSIM can be evaluated by running the predictor on the signals of the

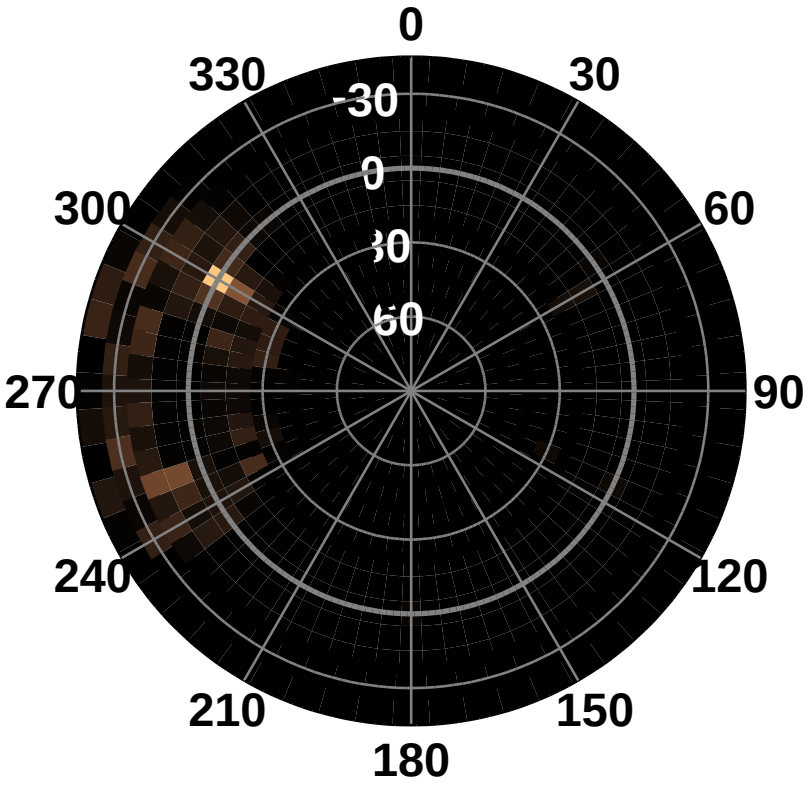
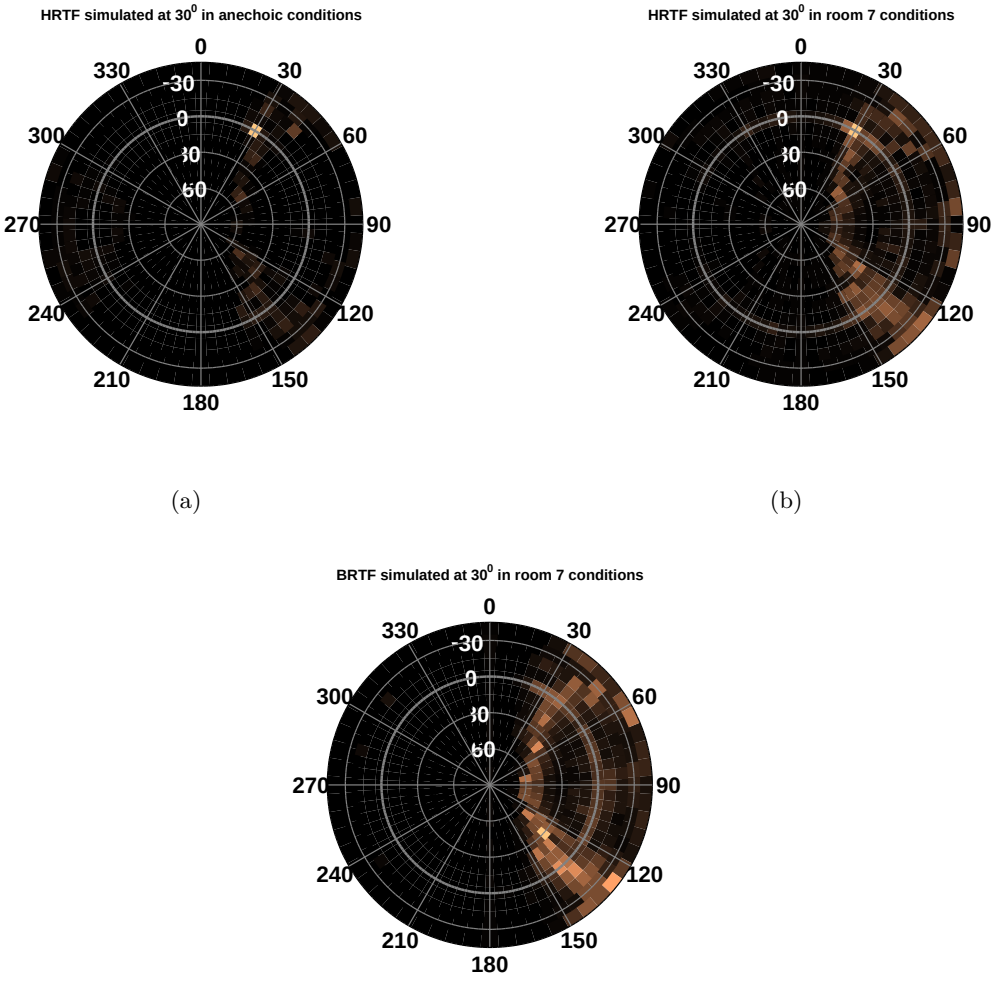


Figure 7.5: *Perceptual representation for a white noise stimulus coming from* 30^o *in anechoic conditions. The black number show azimuth, the white elevation. Light parts show high energy position.*

previous localization experiments. The perceptual maps can be compared with the subjective feedback of the listeners and the test results. The coming sections show comparisons and model predictions for localization, auditory source width and the estimation of front-back confusions. Every of these attributes will be addressed and discussed separately.

7.2.1. Localization prediction

In Fig. 7.6, the following simulations have been carried out. On the left, in Fig. 7.6(a), the input signals to the BASSIM were filtered with HRTFs corresponding to a sound coming from 30^o in anechoic conditions. As expected, the perceptual map shows a compact source, well localized. When adding reverberation to it (Fig. 7.6(b)), the source is still perceived as



(a) (b)

(c)

Figure 7.6: *Effect of reverberation and microphone positions on the perceptual map for positions played at* 30^o*. In (a), the simulation was done in anechoic conditions, in (b) and in (c) in reverberant conditions. In figure (c), the input signals were processed with BTE HRTFs.*

coming from 30º but is now much broader. The perceptual map shows a significant amount of energy across a region spanning azimuths 25º to 50º. The cone of confusion is now much more visible but the position of highest energy is still located at a correct position. The amount of reverberation was set to "room 7", the simulated playback in the localization experiments (see Chapter 4)

In Fig. 7.6(c), the input signals were filtered with the BRTFs measured at 30º. The

acoustical conditions were the same as in Fig. 7.6(b). The effect of the BTE microphone position and the loss of pinna cues result in a much stronger cone of confusion. Most of the energy is present in the back hemisphere and above the horizontal plane (elevation angle of 30°).

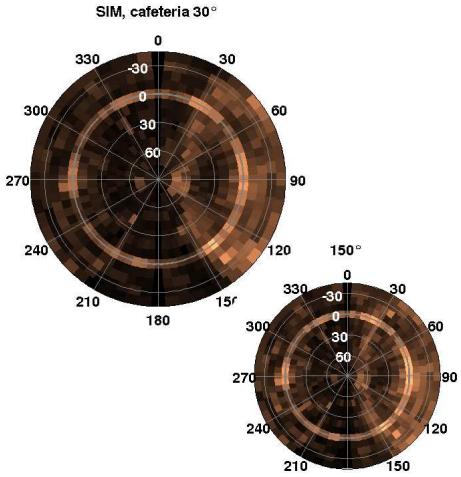
7.2.1.1. Influence of hearing aid algorithms on the perceptual maps

The BASSIM has been developed with the aim of predicting the influence of hearing aid algorithms on spatial perception. In Chapter 4, sound localization with an omnidirectional microphone, a static beamformer and a noise canceler was investigated. Four realistic scenes were reproduced using virtual acoustics. The scene generation procedure was described in details in Chapters 4 and 3. In this section, the cafeteria scene was analyzed with the BASSIM. The scene consists of a male talker speaking in a crowded cafeteria. The diffuse background noise is simulated by twelve positions around the listener. The same signals that were played during the localization experiments were processed by the binaural simulator. The perceptual maps for the three algorithms and the reference (sim condition) that have been generated by BASSIM are shown in Fig. 7.7.

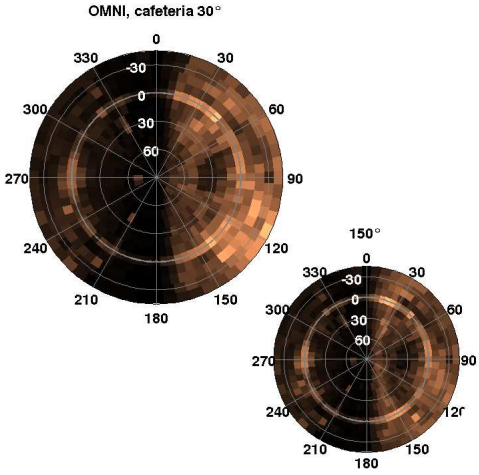
Fig. 7.7 shows the perceptual maps for a target signal played at 30° (large circles). To illustrate the effect of front-back confusions, the perceptual maps for the 150° direction are shown as well (small circles). The sim condition is the reference (Fig. 7.7(a)). Here, the signals were generated with HRTFs measured at the CIC positions. The diffuse background noise is visible on the figure as a high energy circle on the horizontal plane (0° elevation). Recall that the noise was modeled as twelve cardioid sources located on the horizontal plane. They emit mostly towards the receiver. The limited reverberation in the room explains why no noise components are detected at other elevations.

The source components in the perceptual maps of the sim condition can be seen in the graphs as the ellipse that passes by positions 30° and 150° on the horizontal plane across different elevation angles. This is the cone of confusion. Most of the source energy however lays at the correct source position. By looking at the back image (Fig. 7.7(a), small circles), a similar pattern appears. In the subjective listening experiment for this condition, some front-back confusions were observed. This can be seen in the perceptual plots as well.

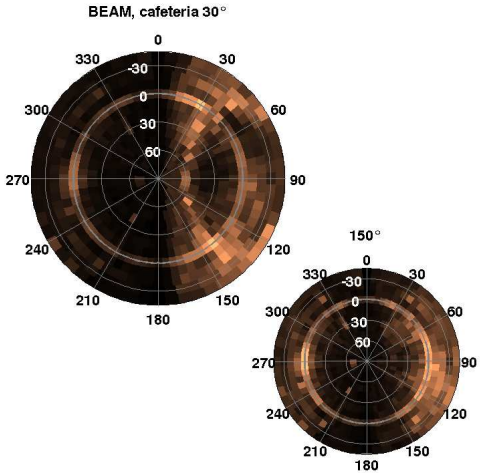
The BTE omnidirectional condition is shown in Fig. 7.7(b). Here, the signals were generated using BRTFs (i.e. HRTFs measured at the BTE microphone positions, see Chap. 4). The BRTFs have been interpolated to other positions on the horizontal plane. For the elevations, a spherical head model has been used to estimate the delays between the four microphones of the BTEs. The amplitudes were interpolated to the corresponding azimuths. The perceptual plots for the omnidirectional condition show as well the background noise on the horizontal plane only. The cone of confusion is present as well. Compared to the sim condition, the cone of confusion appears blurred. It is more diffuse, the positions on the cone are less defined. For the 30° position, more source energy appears to be in the back, close to 120° and at lower elevations. While the perceptual maps for the sim condition offer clear predictions on the position of the target source, the perception associated with the BASSIM output for the omni condition results in a diffuse and uncertain source position. The results



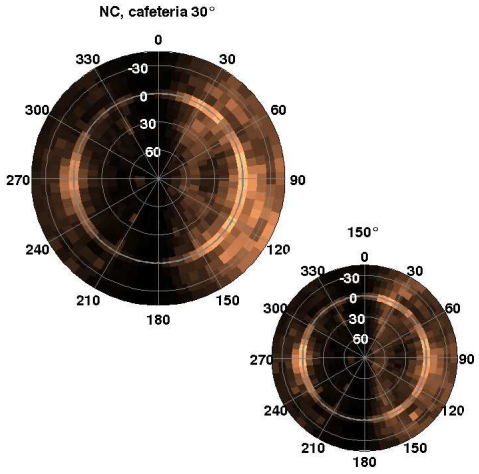
(a) Simulation



(b) BTE omnidirectional



(c) BTE beamformer



(d) BTE noise canceler

Figure 7.7: *Perceptual maps for the three algorithms evaluated in Chapter 4 and the reference condition (sim, Fig. 7.7(a)). The maps were generated using the same signals as in the cafeteria condition.*

of the localization test confirm this impression. For this condition, front-back confusions were at chance level. The rms errors were twice as high as in the sim condition. The rms errors were higher in the front than in the back. The perceptual maps confirm these effects.

The main effect of the beamformer algorithm (Fig. 7.7(c)) is the attenuation of sound sources coming from the back of the listener. By comparing the perceptual maps for the 30° and 150° positions, it can be observed that the components of the target source disappears in the perceptual map for the back position. Only components of the noise are detected, mostly at the sides. For the front position, the cone of confusion is clearly visible. Compared to the sim condition, more source components are perceived in the back. This is due to the loss of

pinna cues. It appears that the binaural cues are better preserved for the beamformer than for the omnidirectional microphone configuration. The source characteristics are therefore better preserved by the beamformer, when its position is in the front of the listener. In the localization experiment, the test subjects had rms errors in the front of the same order than for the sim condition. This is confirmed by the perceptual map analysis. Due to the directivity of the beamformer, the test subjects could resolve all front-back ambiguities because of their a priori expectation on the loudness of the source. A loud source could only come from the front. In comments after the experiments they shared the observation that frontal sources appeared to them in the back as well. Without the loudness expectation they could not separate between front and back positions. This is confirmed by the strong cone of confusion that can be seen in Fig. 7.7(c).

The last algorithm evaluated was the noise canceler (NC). The output of BASSIM for this algorithm is shown in Fig. 7.7(d). For the speech signal played at 30°, the sound source appears very diffuse on the perceptual map. Most of the energy is located on an area around 120° spread across various elevations. The noise appears attenuated compared to the reference condition. Around 0° and 180° it is not visible anymore on BASSIM's output. On the left, at 270°, strong noise components can be seen. The noise attenuation of the left hearing aid was not strong enough to reduce all the noise coming from this direction. Furthermore, due to the difference in SNRs for both hearing aids (due to the position of the source), a strong noise attenuation might have been applied by the right hearing aid. This would have increased the ILDs and moved the noise components to the sides. In the listening test however, no participant mentioned hearing a strong noise at this position.

As in the omnidirectional condition, the source appears more diffuse for a 30° playback position. We relate this to the lack of high frequency pinna cues. For sources played in the front, the pinna cues emphasize high frequencies. In BRTFs, the high frequency content is reduced compare to normal HRTFs.

Globally, the observations done on the perceptual maps coincide with the results of the localization experiment. Each algorithm has a different effect on the perceptual maps.

7.2.2. Source width

As stated in section 2.3, the perceived width of a sound source is related to the interaural coherence and the fluctuations of the interaural cues created by early reflections. The way BASSIM deals with binaural signals with low coherence was tested by feeding the model with a linear combination of two white noise signals.

$$y_l(t) = n_1(t) \star h_l(t) \tag{7.9}$$

$$y_r(t) = (\alpha n_1(t) + (1-\alpha)\, n_2(t)) \star h_r(t) \tag{7.10}$$

where $y_l(t)$ and $y_r(t)$ are the left and right inputs to the BASSIM, $h_l(t)$ and $h_r(t)$ the corresponding HRTFs and $n_1(t)$, $n_2(t)$ two independent white noise signals. The parameter α

is set between 0 and 1 and defines the amout of correlation between the left and right inputs. The $\star$ operator denotes convolution.

The perceptual maps for four different interaural correlation values are shown in Fig. 7.8. Fig. 7.8(d) was generated with α set to 1. It is the same condition than shown in Fig. 7.5 and is the reference situation. The source is compact and well defined positioned 60^o left of the listener. It can be seen in the figures that with reduced interaural coherence the high energy positions on the perceptual maps increase. This implies that perceived width of the sound source increases as well. With an IC close to 0 (Fig. 7.8(a)) the perceptual map shows activity on both left an right hemisphere. When listening to such a signal through headphones, one perceived a very diffuse source inside the head. The perception corresponds to the prediction.

Increasing the interaural coherence reduces the areas of high energy in the perceptual maps (see Fig. 7.8(b) to 7.8(d)). For an interaural coherence of 0.33 it appears that there is still energy on the contralateral side. When the coherence reaches 0.5, the source is clearly localized at 60^o on the left but is much broader than in the reference situation.

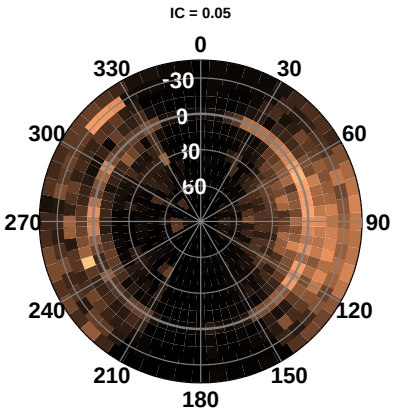
Using Eq. 2.1 to compute the auditory source width gives ASW values of 0.95, 0.97, 0.99 and 1.01 for ICs of 1, 0.5, 0.36 and 0.05 respectively. The ASW computation also shows an increase of auditory source width with reduced coherence. The values are however difficult to interpret and the differences small compared to the perceptual maps. Eq. 2.1 was designed primarily for the analysis of room impulse responses, which might explain the lack of precision and the sensitivity to the input stimulus for this measure.

7.3. Conclusion

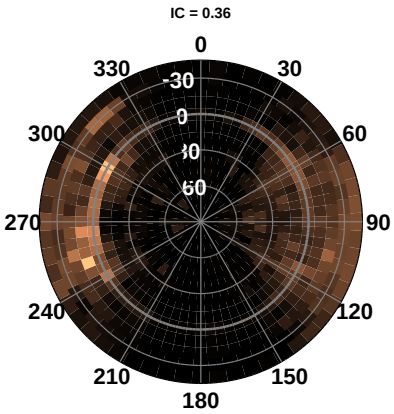
In this chapter, the Binaural Auditory System Simulator (BASSIM) was introduced. This tool was developed in order to automatically assess the spatial quality of hearing aid algorithms. BASSIM is based on Breebaart's model of binaural detection. For any binaural input, it computes an analysis of the spatial information present in the signal in a similar way as done in the human auditory system. Various stages of the human auditory system are taken in consideration in the model. The frequency decomposition that takes place in the cochlea for example was modeled by a set of gammatone filters. The binaural detector that follows extracts at each frequency band and for each frame the best possible ITD-ILD combination. The binaural detector is composed of Excitatory-Inhibitory elements that model the binaural processing of the Lateral Superior Olive (LSO) in the human auditory brainstem(see Chapter 2.5).

A statistical classifier is used to give predictions on the position and width of the input signal. The predictions are displayed in perceptual maps that represent the internal representation a specific subject has of the input signal. The perceptual maps are divided into 710 positions around the subject. At this stage no distance nor internalization perception was modeled by BASSIM.

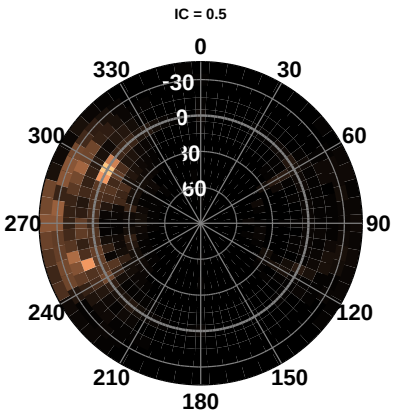
The influence of reverberation and interaural correlation on the output of BASSIM was discussed. It has been shown that reverberation increased the diffusity and the front-back



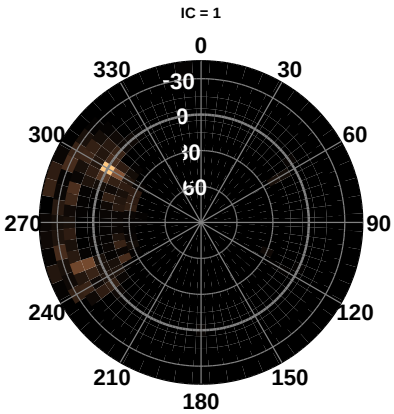
(a) Interaural coherence = 0, $\alpha = 0$



(b) Interaural coherence = 0.36, $\alpha = 0.33$



(c) Interaural coherence = 0.5, $\alpha = 0.5$



(d) Interaural coherence = 1, $\alpha = 0$

Figure 7.8: *Effect of the interaural correlation on spatial prediction for a sound played at* 60^o *in anechoic conditions. Binaural signals with four levels of interaural coherence between 0 and 1 were processed by the BASSIM.*

uncertainty of the input signal. The interaural correlation affects the width of the source as well as predicted by theory. With an correlation close to 0, the energy of the source is spread over all the perceptual maps. This correspond to a very diffuse perception of the sound source.

The BASSIM was used to evaluate the hearing aid algorithms that were tested previously in the localization experiment (see Chap. 4). The signals of the cafeteria scene after hearing aid processing were analyzed by BASSIM. The generated perceptual maps were compared to the ideal reference condition. The observations obtained on the perceptual maps confirm the results of the localization experiment and the feedback comments of the test subjects. The main effects observed were strong front-back confusions and a change in the high energy

regions of the perceptual maps that are consistent with the observed results..

8. Hearing aid algorithm for improved spatial perception

8.1. Introduction

In the majority of today's hearing aids a bilateral audio-processing scheme is implemented, where the individual devices on the left and right ear work independently from each other. The main problem with this type of processing is the loss of the so called binaural information that is primarily responsible for spatial sound perception. This binaural information, that is deduced from signal differences between both ears, does not only enable the listener to precisely locate individual sound sources, but this spatial separation also considerably increases speech intelligibility [Bronkhorst & Plomp 1988]. Recently, hearing aids have been introduced that make use of audio processing schemes where a link between the two devices is incorporated to allow the preservation of binaural information. Current research focuses on refining these existing binaural algorithms [Reindl *et al.* 2010, van den Bogaert 2008] as well as increasing their robustness in an effort to let hearing aid users benefit from the increased speech intelligibility gained by unaltered binaural information. In this chapter we propose a candidate algorithm that increases the perception of a sound source of interest while preserving the auditory space and additionally deals with speech distortion introduced by room reverberation.

Most commercial hearing aids use beamforming algorithms [Hoshuyama *et al.* 1999] to reduce the amount of noise amplified by the system. These can incorporate fixed and adaptive components, where in the fixed version the amplification depends only on the direction of the incoming sound, assuming for instance that the desired sound is always in front of the listener. The adaptive part on the other hand can adjust its directional amplification based on the signal it receives, thereby trying to amplify in the direction of the desired signal and setting noise directions to zero. These algorithms are pretty robust and work well in certain conditions but have major drawbacks: misjudging of the directional setup results in amplified noise and reduced speech signals, noise signals coming from similar directions as the speech signal are equally amplified and since these bilateral algorithms do not preserve the auditory space the listener cannot profit from binaural cues as discussed in the previous chapters, often resulting in perceived sound localized 'in the head'.

With the recent advances of wireless technology, it is possible to use collaborative bilateral hearing systems. Sharing information between the two ears allows the development of improved algorithms that exploit the binaural processing of the human auditory system. Recently, new binaural algorithms have been proposed. Recent research has been carried out on new binaural algorithms that include the Multi-Channel

Wiener Filter (MWF), [van den Bogaert 2008, Doclo & Moonen 2002], binaural beamformers, blind source separation algorithms [Aichner *et al.* 2007] or interaural coherence algorithms [Wittkop 2001, Wittkop & Hohmann 2003].

The MWF approach tries to suppress the corrupting noise from a signal by using the statistical properties of the speech and noise signal components. Combined with a Voice Activity Detector (VAD) the MWF can estimate the optimal Wiener filter. The MWF does not require a priori knowledge of the signals and the microphone positions to deliver an optimal solution in the mean-square sense. Wiener filters in hearing aids have traditionally been monaural algorithms. Until recently [van den Bogaert 2008] they have been made binaural by adding constraints when computing the Wiener filters in order to reproduce correctly the interaural time and level information. It has been shown that this particular algorithm increases localization performance of hearing aid users in simple acoustical conditions.

The blind source separation algorithm proposed by [Aichner *et al.* 2007] adds post-processing adaptive filters to the traditional source demixing techniques. The adaptive filters try to cancel the interfering source components using the estimation parameters of the BSS algorithm. The interference cancellation is applied to a delayed input signal. The delay is the same for the left and right microphones of the hearing aids and the interaural time differences are not distorted. After interference suppression, the resulting signal contains the unprocessed desired signal. The second approach mentioned in the paper applies constraints on the BSS algorithm itself in order to avoid a loss of interaural information in source estimation. However, due to small complexity requirements, the amount of data exchanged between the two devices is limited and the cue distortion problem remains an important issue.

8.2. Algorithm

The algorithm developed in this project combines a binaural noise reduction algorithm with a binaural dereverberation technique that explicitly considers the binaural cues. Combined, both algorithms should reduce the noise components and the effect of reverberation on speech. The noise reduction algorithm is based on the MWF implementation as described in [van den Bogaert 2008, Doclo & Moonen 2002]. The output of the MWF is coupled with the dereverberation algorithm based on the work of [Jeub *et al.* 2010]. The algorithm is described in details in the following sections.

8.2.1. Binaural noise reduction :Multi-channel Wiener filter

The MWF requires a voice-activity detector (VAD) that indicates whether speech is present or not. Many different approaches for such a VAD exist [Ramirez 2004, Marzinik & Kollmeier 2004], but for this project we will just assume having a VAD working with 100% precision. The MWF algorithm works as follows.

Consider the setup depicted in Fig. 8.1, where the speech and noise signals are captured by a left and a right hearing aid with an array of ML and MR microphones respectively. The m-th microphone signal at the left ear can be written in the frequency domain as

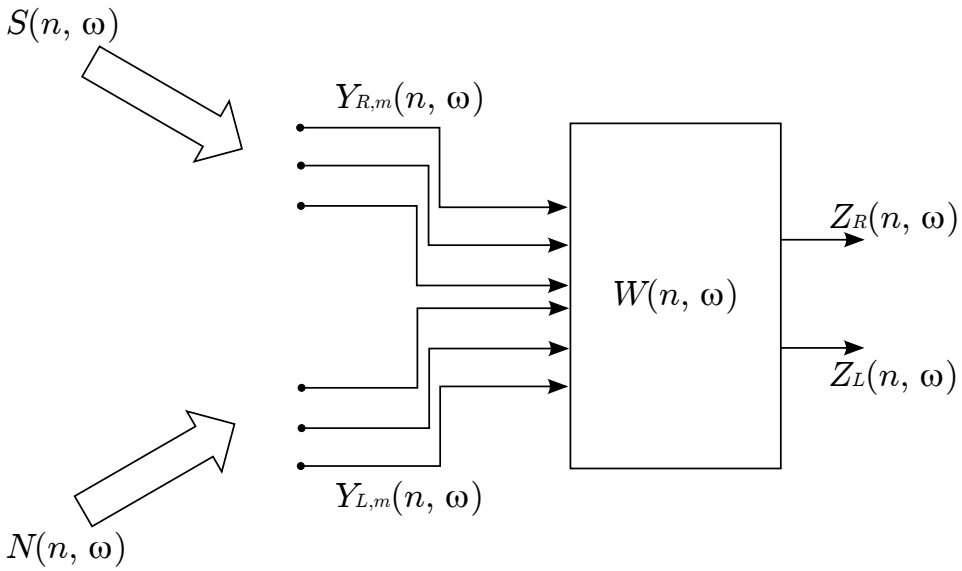


Figure 8.1: *Noise reduction layout with a left and right hearing aid consisting of microphone arrays*

$$Y_{L,m}(n,\omega) = X_{L,m}(n,\omega) + V_{L,m}(n,\omega) \tag{8.1}$$

where $X_{L,m}(n,\omega)$ and $V_{L,m}(n,\omega)$ represent the short time Fourier transforms of the speech and noise components at the m-th microphone, which consist of the speech and noise source signals $S(n,\omega)$ and $N(n,\omega)$ convolved by the respective room impulse response. Assuming that there is a link between the two hearing aids we can use the whole input signal vector $\mathbf{Y}(n,\omega)$ to compute the output signal, where the M-dimensional vector (with $M = M_L + M_R$) $\mathbf{Y}(n,\omega)$ is defined as

$$\mathbf{Y}(\mathbf{n},\omega) = \left[Y_{L,0}(n,\omega)...Y_{L,M_{L-1}}(n,\omega)Y_{R,0}(n,\omega)...Y_{R,M_{R-1}}(n,\omega)\right]^T \tag{8.2}$$

and can be written as

$$\mathbf{Y}(n,\omega) = \mathbf{X}(n,\omega) + \mathbf{V}(n,\omega) \tag{8.3}$$

with $\mathbf{X}(n,\omega)$ and $\mathbf{V}(n,\omega)$ defined similar to $\mathbf{Y}(n,\omega)$. The output signal for the left and right hearing aid $Z_L(n,\omega)$ and $Z_R(n,\omega)$ are obtained by

$$Z_L(n,\omega) = \mathbf{W}_L^H(n,\omega)\mathbf{Y}(n,\omega) \tag{8.4}$$

$$Z_R(n,\omega) = \mathbf{W}_R^H(n,\omega)\mathbf{Y}(n,\omega) \tag{8.5}$$

where $\mathbf{W}_L^H(n,\omega)$ and $\mathbf{W}_R^H(n,\omega)$ are M-dimensional complex vectors representing the left and

right filters with

$$\mathbf{W}(n,\omega) = \begin{bmatrix} \mathbf{W}_L(n,\omega) \\ \mathbf{W}_R(n,\omega) \end{bmatrix} \tag{8.6}$$

The goal of the MWF is to find these vectors $\mathbf{W}_L$ and $\mathbf{W}_R$ so that the cost function

$$J(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_{L,r_L} - \mathbf{W}_L^H \mathbf{Y} \\ X_{R,r_R} - \mathbf{W}_R^H \mathbf{Y} \end{bmatrix} \right\|^2 \right\} \tag{8.7}$$

is minimized. X_{L,r_L} and X_{R,r_R} are the speech components at the left and right reference microphones that the filter tries to estimate and $\mathcal{E}$ is the expected value operator. This equation can be solved by setting the derivative

$$\frac{\partial J(\mathbf{W_L})}{\partial \mathbf{W_L}} = -2\mathcal{E}\{\mathbf{Y} X_{L,r_L}^*\} + 2\mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\}\mathbf{W}_L \tag{8.8}$$

to zero. The optimal multi-dimensional Wiener filter is equal to

$$\mathbf{W}_L = \mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{e}_{L,r_L} \tag{8.9}$$

with $\mathbf{R}_{yy} = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\}$ the $M \times M$ correlation matrices defined as

$$\mathbf{R}_{yy} = \begin{bmatrix} P_{y_0y_0} & P_{y_0y_1} & \cdots & P_{y_0y_{M-1}} \\ P_{y_1y_0} & P_{y_1y_1} & \cdots & P_{y_1y_{M-1}} \\ \vdots & \vdots & \ddots & \vdots \\ P_{y_{M-1}y_0} & P_{y_{M-1}y_1} & \cdots & P_{y_{M-1}y_{M-1}} \end{bmatrix} \tag{8.10}$$

with $P_{y_n y_m}$ the power spectral density or cross power spectral density of the microphone inputs respectively. $\mathbf{R}_{yx} = \mathcal{E}\{\mathbf{Y}\mathbf{X}^H\}$ is defined similarly.

The following two assumptions have to be made to be able to solve the problem. First, it is assumed that the second order statistics of the noise component are sufficiently stationary, so that the estimation for $\mathbf{R}_{vv}$ made during 'noise only' periods can be used during speech periods. The second assumption is that the speech and noise signals are statistically independent, meaning that $\mathbf{R}_{vx} = \mathcal{E}\{\mathbf{V}\mathbf{X}^H\} = 0$

We can then write the optimal filter as

$$\mathbf{W}_L = \mathbf{R}_{yy}^{-1}(\mathbf{R}_{yy} - \mathbf{R}_{vv})\mathbf{e}_{L,r_L} \tag{8.11}$$

where $\mathbf{R}_{yy}$ is estimated during speech and $\mathbf{R}_{vv}$ during 'noise only' periods.

An extension to this filter slightly modifies the cost function by introducing a parameter μ that provides a trade-of between noise reduction and speech distortion. The cost function of this so called speech distortion weighted multichannel Wiener filter (SDW-MWF) can be written as:

$$J(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_L - \mathbf{W}_L\mathbf{X} \\ X_R - \mathbf{W}_R\mathbf{X} \end{bmatrix}\right\|^2 + \mu \left\|\begin{bmatrix} \mathbf{W}_L\mathbf{V} \\ \mathbf{W}_R\mathbf{V} \end{bmatrix}\right\|\right\} \tag{8.12}$$

where the first term represents speech distortion energy and the second one the residual noise component. For $\mu = 1$ the SDW-MWF corresponds to the standard MWF, if $\mu > 1$ the filter focuses on noise reduction at the cost of speech distortion and for $\mu < 1$ more emphasis is put on speech preservation at the expense of less noise reduction. The optimal filter can then be written as

$$\mathbf{W}_L = (\mathbf{R}_{yy} + (\mu - 1)\mathbf{R}_{vv})^{-1}(\mathbf{R}_{yy} - \mathbf{R}_{vv})\mathbf{e}_{L,r_L} \tag{8.13}$$

Preservation of cues To evaluate the preservation of the acoustic scene the interaural transfer function is introduced, which fully describes the interaural cues. The input and output ITF are defined as the ratios of the signal components at the left and right ear.

$$\begin{aligned} ITF_x^{in} &= \frac{X_{L,r_L}}{X_{R,r_R}}, \quad ITF_x^{out} = \frac{\mathbf{W}_L^H\mathbf{X}}{\mathbf{W}_R^H\mathbf{X}} \\ ITF_v^{in} &= \frac{V_{L,r_L}}{V_{R,r_R}}, \quad ITF_v^{out} = \frac{\mathbf{W}_L^H\mathbf{V}}{\mathbf{W}_R^H\mathbf{V}} \end{aligned} \tag{8.14}$$

Maintaining the same ITF for every frame at the output as at the input implies having the same binaural information after applying the filter as without it. A detailed analysis of the binaural MWF as it is done in [van den Bogaert 2008] chapter 4.3.2 shows that the MWF vectors $\mathbf{W}_L$ and $\mathbf{W}_R$ are parallel, with

$$\mathbf{W}_L = \alpha\mathbf{W}_R \tag{8.15}$$

where $\alpha^* = \mathbf{X}_L \div \mathbf{X}_R = ITF_{in}$ is the complex conjugate of the ITF of the speech component at the input. The ITF of the output speech and noise components reduce to ITF_x^{in} and ITF_v^{in} respectively.

$$\begin{aligned} ITF_x^{in} &= \frac{\mathbf{W}_L^H\mathbf{X}}{\mathbf{W}_R^H\mathbf{X}} = ITF_x^{out} \\ ITF_v^{in} &= \frac{\mathbf{W}_L^H\mathbf{V}}{\mathbf{W}_R^H\mathbf{V}} = ITF_v^{out} \end{aligned} \tag{8.16}$$

This means, that only the binaural cues of the speech component are preserved and that the noise will be perceived as also coming from the direction of the speech.

8.2.2. Binaural dereveberation

To get rid of reverberation effects some algorithms use harmonic filtering where the fact, that speech is composed of many overlaying harmonic waves originating from the vocal chords, is utilized. This approach however can only be used when the desired signal is composed of speech and even though this might be a common case for hearing aids it is a severe restriction.

Spectral subtraction The binaural dereverberation method used in this project was developed by [Jeub *et al.* 2010] and consists of two consecutive parts. The first stage aims at getting rid of late reverberation effects that arrive at the listener with a time delay larger than $T_l = 100ms$. The reverberant signal $x(k)$, which is the convolution of the clean signal $s(k)$ with the room impulse response (RIR) $h(k)$ of length T, can be divided into early and late reverberation components

$$x(k) = x_{early}(k) + x_{late}(k) = \sum_{m=0}^{T_l f_s - 1} s(k-m)h(m) + \sum_{nm=T_l f_s}^{Tfs} s(k-m)h(m) \tag{8.17}$$

with the sampling frequency f_s. The late reverberant speech is viewed as an uncorrelated noise process that is obtained from a simple model of the late part of the RIR $h_{late}(k)$

$$h(k) = h_{late}(k) = n(k)e^{-\rho k \div f_s}; \quad for T_l f_s \leq k \leq T f_s \tag{8.18}$$

where $n(k)$ is a sequence of normally distributed random variables with zero mean and $\rho = 3ln(10)/divRT60$ a function of the reverberation time. The variance of the late reverberant speech can now be estimated by

$$\sigma^2_{x_{late}}(n,\omega) = e^{-2\rho T_l}\sigma^2_x(n - N_l, \omega) \tag{8.19}$$

where σ^2_x is the variance of the reverberant speech signal and N_l the number of STFT frames in T_l. Then the signal to interference ratio corresponds to:

$$SIR(n,\omega) = \frac{\mid X(n,\omega) \mid^2}{\sigma^2_{x_{late}}} \tag{8.20}$$

and the weights for the filtering of the late reverberation effects are computed to induce a spectral magnitude subtraction.

$$G_{late}(n,\omega) = 1 - \frac{1}{\sqrt{SIR(n,\omega)}} \tag{8.21}$$

These weights are actually only calculated once using a reference signal composed of the reverberant input of the left and right hearing aid and are then applied to both signals alike.

Sound field coherence filter After suppressing the late reverberation components the second part of the procedure tries to eliminate early reflections that arrive very shortly after the direct speech. This is done by utilizing the fact that the direct components have a very high coherence, whereas the early reverberation can be modeled as additive uncorrelated noise sources with a low coherence based on the sound field. The filter aims to remove non-coherent signal parts while keeping the coherent parts intact. As in the previous sections, we will decompose the reverberant signal into direct speech and early reverberation based on the time delay of arrival with the threshold set to $T_d = 2$ms.

$$x(k) = x_{direct}(k) + x_{early}(k) = \sum_{m=0}^{T_d f_s - 1} s(k-m)h(m) + \sum_{m=T_d f_s}^{XTf_s} s(k-m)h(m) \tag{8.22}$$

We now use the the minimum squared error criterion $\|S - G_{\text{coh}}\|^2$ to determine the optimal filter weights. Solving this yields:

$$G_{\text{coh}}(n,\omega) = \frac{\Phi_{ss}(n,\omega)}{\Phi_{ss}(n,\omega) + \Phi_{nn}(n,\omega)} \tag{8.23}$$

where $\Phi_{ss}(n,\omega)$ and $\Phi_{nn}(n,\omega)$ denote the auto power spectral density of the undisturbed speech signal and the additive reverberation component respectively.

To compute these weights we consider having time aligned signals that have the same reverberation PSD on both the left and right hearing aid. We can then decompose the actual measured PSD as follows:

$$\begin{aligned} \Phi_{x_l x_l}(n,\omega) &= \Phi_{ss}(n,\omega) + \Phi_{nn}(n,\omega) \\ \Phi_{x_r x_r}(n,\omega) &= \Phi_{ss}(n,\omega) + \Phi_{nn}(n,\omega) \\ \Phi_{x_l x_r}(n,\omega) &= \Phi_{ss}(n,\omega) + \Gamma_{x_l x_r}(\Omega)\Phi_{nn}(n,\omega) \end{aligned} \tag{8.24}$$

where $\Gamma_{x_l x_r}(\Omega)$ is the soundfield coherence between the two hearing aids. The resulting weights can now be calculated by:

$$G_{\text{coh}}(n,\omega) = \frac{\hat{\Phi}_{ss}(n,\omega)}{\frac{1}{2}\left(\hat{\Phi}_{x_l x_l}(n,\omega) + \hat{\Phi}_{x_r x_r}(n,\omega)\right)} \tag{8.25}$$

where the PSD estimates $\hat{\Phi}_{ss}(n,\omega)$, $\hat{\Phi}_{x_l x_l}(n,\omega)$ and $\hat{\Phi}_{x_r x_r}(n,\omega)$ are computed as shown in the next section.

Preservation of cues Since both left and right signal are filtered with the exact same coefficients in both parts of the dereverberation process the ITF and the binaural cues in every frame also stay the same. In [Raspaud *et al.* 2010] it can be seen, that preserving the per frame binaural cues is sufficient for source localization and therefore the dereverberation should not alter the acoustic scene.

8.3. Implementation

In the implementation setup the SDW-MWF has to be supplied with an input signal for every simulated hearing aid microphone together with the corresponding VAD. The SDW-MWF then carries out the noise reduction and sends a left and a right output signal to the dereverberation stage. There all calculations are based on the time aligned versions of the inputs and the identical filter is applied to both sides.

8.3.1. Multi-channel Wiener filter

For the actual implementation of the MWF only one microphone was simulated at each hearing aid, making it easier to create sensible input data and reducing the overall complexity. The frame-wise updating equations for the noise and signal correlation matrices then reduce to:

$$\begin{aligned} \mathbf{R}_{yy}(n,\omega) &= \alpha_1 \mathbf{R}_{yy}(n-1,\omega) + (1-\alpha_1) \begin{bmatrix} P_{y_l y_l}(n,\omega) & P_{y_r y_l}(n,\omega) \\ P_{y_l y_r}(n,\omega) & P_{y_r y_r}(n,\omega) \end{bmatrix} \\ \mathbf{R}_{vv}(n,\omega) &= \alpha_2 \mathbf{R}_{vv}(n-1,\omega) + (1-\alpha_2) \begin{bmatrix} P_{v_l v_l}(n,\omega) & P_{v_r v_l}(n,\omega) \\ P_{v_l v_r}(n,\omega) & P_{v_r v_r}(n,\omega) \end{bmatrix} \end{aligned} \tag{8.26}$$

where α_1 and α_2 are smoothing factors and the individual power spectral densities are calculated with the Welch method. The vectors e_{L,r_L} and e_{R,r_R} are $[10]^T$ and $[01]^T$ respectively and the filtered output signals are transformed back into the time domain with the overlap-and-add method.

8.3.2. Spectral filter for late reverberation

As mentioned above a time aligned reference signal is used to calculate the late reverberation coefficients. The left and right output signals of the MWF are time aligned using the generalized cross-correlation with phase transform (GCC-PHAT) method [Knapp & Carter 1976] and the reference signal is then computed according to:

$$X_{ref}(n,\omega) = \frac{1}{2}\left(X'_l(n,\omega) + X'_r(n,\omega)\right) \tag{8.27}$$

where the apostrophe denotes time aligned signals. The variance of the reverberated speech signal $\sigma^2_{X_{ref}}(n,\omega)$ required to calculate the SIR is estimated by:

$$\sigma^2_{X_{ref}}(n,\omega) = \alpha_3 \sigma^2_{X_{ref}}(n-1,\omega) + (1-\alpha_3)|X_{ref}(n,\omega)^2| \tag{8.28}$$

with smoothing factor $\alpha 3$. After computing the weights $Glate(n,\omega)$ a lower bound G^{min}_{late} is applied to get rid of overestimation of the variances. A common problem in acoustic filters is musical noise that is introduced by rapidly changing filter coefficients. Therefore a smoothing of the weights is performed by applying a moving average window with a variable length depending on the SIR of the respective frame. The power ratio of untreated to filtered signal is calculated for every frame according to:

$$\zeta(n) = \frac{\sum_\omega |G_{late}(n,\omega) X_{ref}(n,\omega)|^2}{\sum_\omega |X_{ref}(n,\omega|^2} \tag{8.29}$$

and the window length is set to:

$$L_w(n) = \begin{cases} 1, & \text{for } \zeta(n) \geq \zeta_{thr} \\ 2\text{round}\left[\left(1 - \frac{\zeta(n)}{\zeta_{thr}}\right)\psi\right] + 1, & \text{for } \zeta(n) < \zeta_{thr} \end{cases} \tag{8.30}$$

were ζ_{thr} is the threshold between high and low SIR regions and is a weighting factor. This way the lower the SIR of a processed frame is, the larger the window length will become resulting in a stronger smoothing effect. Afterwards the smoothed filter coefficients are applied to the reverberated inputs:

$$\begin{aligned} S_l(n,\omega) &= G_{late}(n,\omega)X_l(n,\omega) \\ S_r(n,\omega) &= G_{late}(n,\omega)X_r(n,\omega) \end{aligned} \tag{8.31}$$

8.3.3. Coherence filter for early reverberation

To compute the weights of the coherence based filter the power spectral density (PSD) estimates for the clean and the reverberant signals are required. The PSD of the clean part can be calculated by:

$$\hat{\Phi}_{ss}(n,\omega) = \frac{\text{Re}\left\{\hat{\Phi}_{x_l x_r}(n,\omega)\right\} - \frac{1}{2}\text{Re}\left\{\Gamma_{x_l x_r}(\Omega)\right\}\left(\hat{\Phi}_{x_l x_l}(n,\omega) + \hat{\Phi}_{x_r x_r}(n,\omega)\right)}{1 - \text{Re}\left\{\Gamma_{x_l x_r}(\Omega)\right\}} \tag{8.32}$$

where again the estimates $\hat{\Phi}_{x_l x_l}(n,\omega)$ and $\hat{\Phi}_{x_r x_l r}(n,\omega))$ for the PSD of the left and right signal are used, as well as the cross power spectral density (CPSD) $\hat{\Phi}_{x_l x_r}(n,\omega)$. These estimates are updated for every frame and computed as follows:

$$\begin{aligned} \hat{\Phi}_{x_l x_l}(n,\omega) &= \alpha_4 \hat{\Phi}_{x_l x_l}(n-1,\omega) + |X'_l(n,\omega)|^2 \\ \hat{\Phi}_{x_r x_r}(n,\omega) &= \alpha_4 \hat{\Phi}_{x_r x_r}(n-1,\omega) + |X'_r(n,\omega)|^2 \\ \hat{\Phi}_{x_l x_r}(n,\omega) &= \alpha_4 \hat{\Phi}_{x_l x_r}(n-1,\omega) + X'_l(n,\omega)X'^*_r(n,\omega) \end{aligned} \tag{8.33}$$

where X'_l and X'_r stand for time aligned signals using GCC-PHAT and α_3 is a smoothing factor. The coherence $\Gamma_{x_l x_r}(\Omega)$ is deduced from the soundfield model that can be applied to the acoustic situation. A simple approach is considering the two hearing aids as two microphones in a spherically isotropic (diffuse) soundfield. The coherence can then be expressed as:

$$\Gamma_{x_l x_r}(\Omega) = sinc\left(\frac{2\pi f d_{mic}}{c}\right) \tag{8.34}$$

with d_{mic} the microphone distance and c the speed of sound. A lower bound is applied to the computed values for $G_{coh}(n,\omega)$ as it was done in section 3.2 and to counter musical noise the coefficients are smoothed with a constant moving average window of length ϕ.

8.4. Evaluation

8.4.1. Simulation setup

In order to evaluate the effect of the implemented algorithms simulations were run over a set of testing parameters. The general setup consists of one speech source located in front of the listener at a distance of 2 meters. The noise source is placed at the same distance with an angular position relative to the speech source ranging from 0 to 180 deg (in steps of 30 deg). The noise source starts sending a modulated noise signal consisting of icra noise [Dreschler *et al.* 2001] and after about 5s the speech source sends a sample sentence from the OLSA database. This experiment was carried out with initial signal-to-noise ratios (SNR) of 0dB and 3dB and in 3 different room setups calculated with the room acoustics simulator ROOMSIM, with respective reverberation times of 0.5s, 1s and 1.5s. The filter parameters used in the simulation are given in Table 8.1 below.

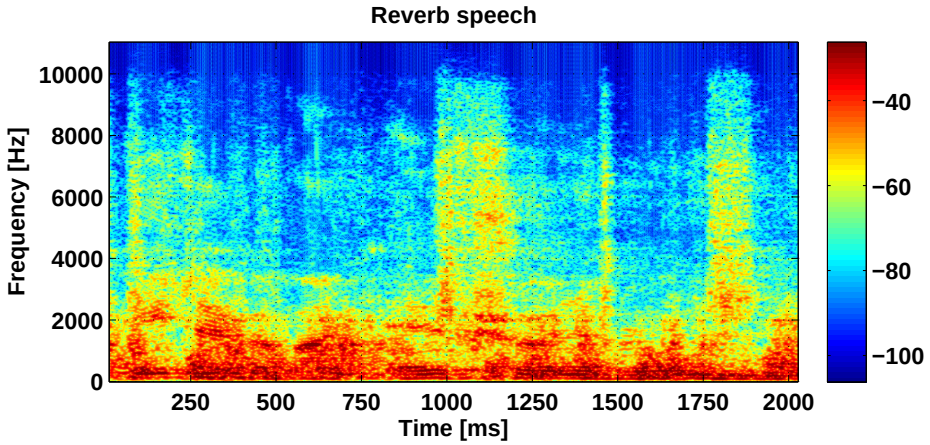
Parameter	Value
Sampling frequency	$f_s = 22050\text{Hz}$
Frame length	$L_F = 512$
FFT length	$L_{FFT} = 512$
Frame overlap	50%
STFT window type	Hann window
Smoothing factors	$\alpha_1 = 0.95, \alpha_2 = 0.99, \alpha_3 = 0.9, \alpha_4 = 0.9$
Speech distortion weighting	$\mu = 5$
Reverberant gain factor thresholds	$G_{late}^{min} = G_{coh}^{min} = 0.3$
Late reveberation time-span	$T_l = 0.1\text{s}$
Late gain smoothing threshold	$\zeta_{thr} = 0.5$
Late gain smoothing scaling factor	$\psi = 25$
Microphone distance	$d = 0.17\text{m}$
Speed of sound	$c = 343.2\text{m/s}$
Coherence gain smoothing length	$\phi = 7$

Table 8.1: *Simulation settings*

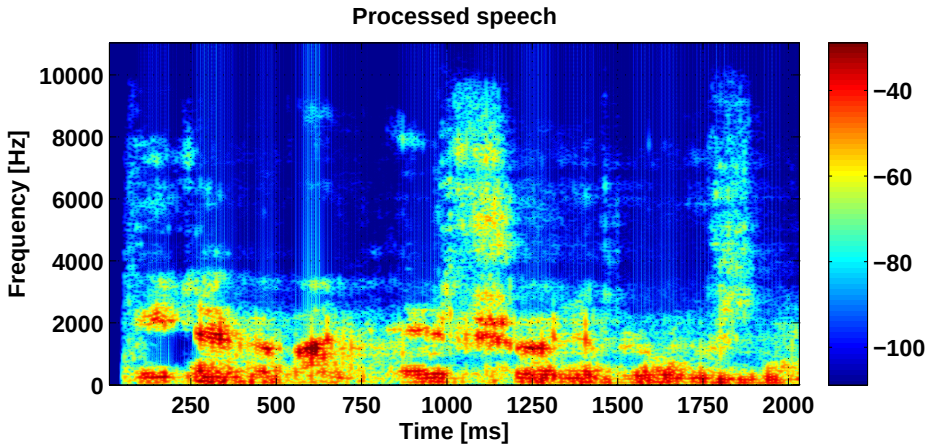
The effect of the dereverberation algorithm on a clean speech signal is illustrated in Fig. 8.2. For a reverberation time of $T_{60} = 1.5\text{s}$ and with the parameters described in Table 8.1, the algorithm effectively suppress the reverberation from the input speech signal.

8.4.2. Objective measures of speech intelligibility

The performance of the noise reduction of the MWF was quantified using the binaural SII (speech intelligibility index) measurement software, developed in the Hearcom project [Beutelmann & Brand 2006]. This is an extension of the monoaural SII [ANSI-SII 1997] also including the 'best ear' effect, as well as the binaural processing of the auditory system (based on the binaural equalization-cancelation model of Durlach [Durlach 1963]). Speech and noise



(a)



(b)

Figure 8.2: *Spectrograms of: a) Reverberant ($T_{60} = 1.5s$) and b) processed speech signals.*

components were analyzed before and after applying the MWF by generating random noise with identical PSD and then calculating the respective speech reception threshold (SRT) values. The SRT describes the minimum SNR required to understand 50% of spoken words correctly, hence a decrease in SRT depicts an increase in speech intelligibility.

Fig. 8.3 shows the SRT values for initial SNR values of 0dB and 3dB respectively. The SRT was measured in all three room setups and then averaged. In both cases a decrease in SRT ranging between 2dB and 5dB was achieved by applying the MWF, indicating a significant increase in intelligibility through the reduction of noise. The noise angle changes the SRT for both the unfiltered and processed case due to the spatial release from masking discussed in section 2.1.

The angle also has a slight influence on the amount of intelligibility increase, as it is smallest for noise coming from the front or the back, implying that the filter profits from the higher

initial SNR in the cases were a 'best ear' exists. To measure the intelligibility increase due to the dereverberation an attempt was made to integrate the speech transmission index (STI) as well as the speech to reverberation modulation energy ratio (SRMR) [Falk & Chan 2008] in the evaluation. However this proved to be more challenging than anticipated and would have gone beyond the scope of this project. Subjective listening tests support the notion that the algorithm does in fact reduce reverberation effects as can also be seen in the spectral energies displayed in Fig. 8.2. But as previously stated the impact of the filter on speech intelligibility could not be quantitatively shown. The theoretical derivation concerning the preservation of cues, was supported by informal listening tests, where the original speech direction was indeed conserved.

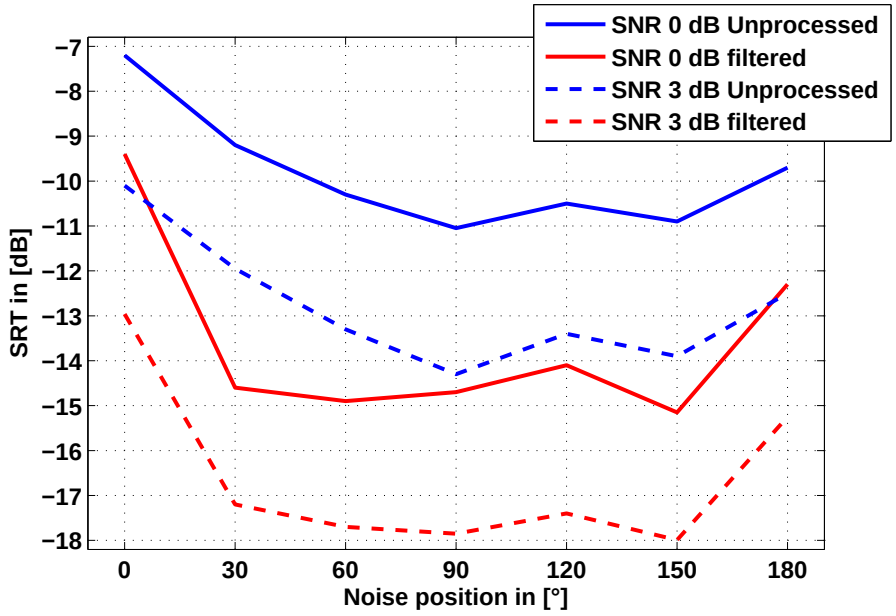


Figure 8.3: *Predicted SII gains for the noise reduction filter (red) against the unprocessed condition (blue) depending on the position of the noise at two SNRs: 0 dB and 3 dB (dashed lines)*

8.4.3. BASSIM prediction

The BASSIM was applied on signals processed by the algorithm. For the sake of comparison, the cafeteria scene was simulated. The algorithm assumed a perfect VAD as was done previously. The outcome of BASSIM's prediction is shown in Fig. 8.4. The analysis is done as in Chapter 7 were the perceptual maps of the omnidirectional, beamformer and noise canceler for the cafeteria were discussed. The situation with a target speaker at 30° was analyzed. This position was chosen as the listeners had difficulties in distinguishing a source in the front from one in the back for signals coming from this direction. The perceptual map for the back

target position (150°) is shown in Fig. 8.4 as well (lower right corner).

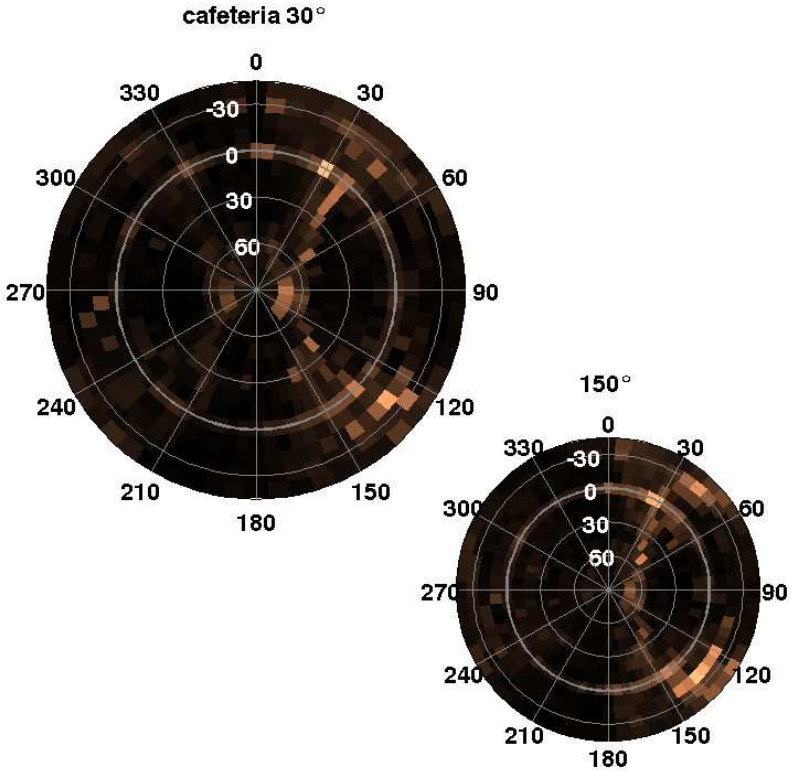


Figure 8.4: *Perceptual maps of the proposed binaural algorithm in the cafeteria condition. The input signals were generated following the procedure described in Chapter 4. The perceptual maps for target positions* 30^o *(large) and* 150^o *are shown.*

Compared to the perceptual maps displayed in Fig. 7.7 the algorithm does a good job at suppressing the diffuse background noise. The ideal condition (perfect VAD, ideal data link) in which the algorithm has been implemented clearly give it an advantage compared to the other. By looking at the spatial representation of the target source, one can notice that for the 30° the spatial characteristics of the target speaker appear relatively more accurate. It is correctly localized at 30° and fairly compact. The cone of confusions is however still visible. It seems slightly off-position, closer to 30°. There is however less energy in the back, which compare favorably to the other algorithms.

For a target signal played at 150° (Fig. 8.4, small plot), a relatively diffuse source appears at around 135°. The excitation is spread over almost thirty degrees and at a lower elevation ($-20°$). Localization judgments are primarily based on spectral pinna cues. For signal coming from behind the listeners, the shape pinna attenuates already the high frequencies. With noise, the spectral contours used by the human auditory system for making elevation decisions could have been corrupted. The MWF and the dereveberation algorithms aim at reproducing the interaural time and level differences accurately. They cannot deal with monaural spectral cues

as this would require a priori knowledge on the spectral profile of the target signal. Here it appears that this information is degraded, which might produce this diffuseness. Nevertheless, the interaural information seems to have been reproduced correctly as the highlighted position lays on the cone of confusion of the $30^{o} - 150^{o}$ source pair. There is still significant energy in the front, indicating that for this position front-back confusions are likely.

8.5. Conclusion

The implemented algorithm performs well in the evaluated environment, the noise reduction significantly increases speech intelligibility. The amount of reverberation can also be reduced with this combination of methods, but the positive effect on speech perception could not be quantified due to a lack of validated evaluation measures. However it should be noted, that the algorithms require prior information about the environment they are working in. The MWF relies on a good voice activity detection, which not only means having less effective filters due to activation detection errors, but also that the application range is limited to setups, where such a VAD can be implemented in a reasonable way. The dereverberation on the other hand depends on the reverberation time of the acoustic scene and even though methods for estimating this parameter exist (for an overview consult [Ramirez *et al.* 2007]) estimation errors could decrease performance.

Furthermore a data link was assumed to transfer all the information of one hearing aid to the other, which results in a lot of data traffic considering multiple microphones at each side. In the ideal case of having a wireless link between both devices the amount of data needed exceeds present-day transmission capabilities and additional noise or delays might be introduced to the system. Future work could test the discussed processing scheme in a more realistic setup, where the algorithm first has to estimate the acoustic characteristics, namely the reverberation time, as well as work with an actual VAD. Additionally, instead of using artificial measurement methods to determine the performance a series of listening experiments with actual test subjects could be carried out. Within the scope of such experiments some localization tests could also be done to analyze the theoretical preservation of binaural information and the effect on spatial perception. This being an active research field with significant application areas in hearing aids and other sound processing systems, improvements, combinations and adaptation of the stated and other algorithms are continuously developed. Promising extensions to the basic SDW-MWF described here combine it with a beamforming preprocessor that makes use of the existing and functioning hearing aid technology [Spriet *et al.* 2004].

9. Conclusions

The main aim of the thesis was to propose new methods that allow the evaluation of hearing instruments in realistic acoustical conditions. This was motivated by the finding that the reproduction of spatial acoustical features by bilateral hearing aids was poorly rated by the users. It has been argued that the signal processing strategies in hearing instruments modify significantly the cues used by the human auditory system to characterize sounds with a position, a distance, a width, etc... Hearing aid algorithms have been primarily designed to attenuate surrounding background noise and increase speech understanding. It is until recently that new algorithms have come to market, that explicitly consider binaural cues. There is however a lack of methods or tools that allow the evaluation of the spatial quality of hearing instruments.

To solve this issue two different approaches have been implemented. The first relies on an efficient and perceptually accurate reproduction of virtual acoustical scenes in which hearing aid algorithms can be perceptually evaluated. The second approach is based on a model of the human binaural auditory system. It offers predictions on the position, width and front-back uncertainty of sound signals processed by hearing aids.

Both methods have been applied to a selection of hearing aid algorithms. The results indicate that the selected algorithm degrade spatial sound perception. The listeners encountered big difficulties to distinguish between sounds played in the front from the back. This confirms the findings of previous studies on sound localization with bilateral hearing aids.

Finally, a new algorithm was introduced. The algorithm combined aspects of the binaural Multichannel Wiener Filter (MWF) [van den Bogaert 2008] with a binaural dereverberation technique [Jeub *et al.* 2010]. It reproduced interaural time and level differences with low distortion while reducing noise and reverberation and thus localization was preserved.

9.1. Overview of achievements

At the beginning of this work, three main objectives were proposed. The work done towards the completion of each objective will be discussed in the following sections. Suggestions for future work and improvements will be discussed as well.

9.1.1. Tools for the evaluation of hearing aid algorithms in realistic conditions

9.1.1.1. System for virtual acoustics

The first tool introduced in this thesis was the system for virtual acoustics (Chapter 3). The system allows the reproduction of complex acoustical environments by combining individual Head-Related Transfer (HRTFs) measurements, room simulations and an accurate and efficient reproduction of head movements. The system was evaluated perceptually in Chapters 3 and 4.

The evaluations have demonstrated that the system was able to propose virtual realities that were perceptually very close to the real world. In some conditions, no difference between virtual and real playbacks could be detected. The fact that the system relied on open transducers for the reproduction of sound reduced the internalization phenomenon observed with other virtual sound reproduction methods. The system was designed to require relatively low computational power and can be run with MATLAB on standard PCs without any problems.

9.1.1.2. Binaural auditory system simulator

The system for virtual acoustics can be used for subjective listening experiments. It has been completed with the Binaural Auditory System Simulator (BASSIM) that allows an automatic evaluation of hearing aid algorithms (Chapter 7). For an arbitrary binaural input signal, BASSIM offers a prediction on the position and the width of the perceived source. The BASSIM implementation presented in this thesis follows closely Breebaart's binaural model. It is composed of a peripherical model and a binaural processor and simulates the processing of the outer, middle and inner ears. The binaural processor is composed of Excitatory-Inhibitory (EI) elements tuned to a specific combination of Interaural Time and Level Differences (ITDs-ILDs). It has been extended with a random forest classifier trained on individual HRTFs that cover 710 positions in space. The input to the classifier is the ITD and ILD combination (the variables τ and α in the model) that produces the minimal response in the binaural processor.

BASSIM has been applied on various acoustical conditions. The impact of Interaural Coherence (IC) and room reverberation on the perceptual prediction has been discussed. It has been shown that a decrease in IC results in a larger high energy area in the perceptual maps. This was interpreted as resulting in the perception of a broader sound source. This confirms classical room acoustics theory. Reverberation increased the perceived widths of sound sources and front-back uncertainty. The cone of confusion (i.e. regions of equal ITDs and ILDs) was stronger marked when reverberation was added to the binaural signals.

9.1.2. Performance of hearing aid algorithms in realistic environments

Three commonly used hearing aid algorithms have been evaluated in this project: an omnidirectional microphone of a Behind-the-Ear (BTE) hearing aid, a first order differential static beamformer and a classic noise canceler (Chapters 4 and 6). The system for virtual acoustics was used to simulate four realistic acoustical environments: a cafeteria, an office, a street

and a forest in which the following signals had to be localized: male speech, a phone, an ambulance siren and a bird.

The results have shown that the selected hearing aid algorithms reduce localization. Due to the positions of the microphones, the listeners have experienced great difficulties to distinguish between sounds played in the front and in the back. This was caused by the loss of the pinna cues. The beamformer, due to its directivity characteristics, allowed the test subjects to resolve the front-back confusions. In the frontal hemisphere, performance for the beamformer was close to the reference condition in the office, street and forest scenes. As the cafeteria was the only condition where the target signal had significant low frequency energy, these results suggest that the ITDs were not reproduced accurately by the algorithm.

Five hearing impaired subjects with symmetrical hearing loss participated in the localization experiment. For these subjects, the conditions tested were the cafeteria and the office with the omnidirectional and beamformer algorithms. The results show a degradation of localization performance compared to the normal listeners for the same conditions. For these listeners as well, the beamformer removed almost all front-back confusions.

In Chapter 6 an experiment was implemented in which the perception of sound source distance was investigated. The same algorithms as in the localization experiment were tested. Based on the assumption that the hearing aid algorithms modified the main distance cues (sound intensity and direct-to-reverberation ratio) a degradation in distance perception was expected. However, we did not find this effect in the results. The strong variation in the subjects' response might explain why the test was inconclusive.

Head movements and localization with bilateral Cochlear Implants (CI) was investigated in Chapter 5. The listeners had to localize speech signals of different durations in background noise. Two test conditions were investigated. In the first, the listeners had to keep their head still, fixing the loudspeaker in front of them. In the second condition, they were allowed to move their heads freely on the horizontal plane. The experiment illustrated that head movements are essential for bilateral CI users to resolve front-back confusions. These findings suggest that in order to quantify the real gain CI users get from their devices, experiments need to be carried out in situations that are close to their daily environment. The speech understanding improvements would then probably be larger than in the artificial clinical environment as well.

9.1.3. Hearing aid algorithm for improved spatial perception

The algorithm discussed in Chapter 8 preserves the spatial cues (essentially interaural time and level differences) while reducing the amount of reverberation and background noise. The Multichannel Wiener Filter noise reduction algorithm was based on previous work by [van den Bogaert 2008]. By adding an extra cost function to the computation of the optimal Wiener filter, this methods explicitly takes into account the interaural cues.

The MWF has been completed by the dereverberation algorithm proposed by [Jeub *et al.* 2010]. This technique divides the signal into early and late reflection parts. The late reflections are defined as arriving 100ms after the direct sound component. They are

modeled as a random process with exponential decay characterized by the reverberation time of the environment. The algorithm decreases the influence of the late reverberation by reducing the Signal-to-Interference Ratio. The SIR is defined as the ratio of the signal power to the spectral variance of the late reflection model. Additionally, this algorithm estimates the running soundfield coherence between the microphones. The coherence value gives an estimate of the amount of early reflections still present in the signal. The uncoherent components of the signal are removed using the filter defined in Eq. 8.25. The interaural cues are respected by this algorithm as identical gain and phase values are applied to the left and right output signals.

The algorithm was evaluated using the binaural Speech Intelligibility Index (SII) in simulated rooms and with the BASSIM. A perfect voice activity detector was assumed. The SII analysis showed an average improvement in the Speech Reception Threshold (SRT) of 4 dB compared to the unprocessed condition. The BASSIM evaluation confirmed that localization was preserved, compared to the other algorithms evaluated in Chapter 4.

9.2. Suggestions for improvement and future work

9.2.1. System for virtual acoustics

The system for virtual acoustics was able to reproduce efficiently a limited number of sound sources. Improving the efficiency of the system or implementing the processing in a real-time platform could increase performance and the complexity of the scenes presented. Another limitation of the simulator is that it relies on the offline computation of room impulse responses. Due to the complexity of the room simulation procedure, the test scenarios have to be defined prior to testing.

The system relies on individual HRTF measurements. In this project, the HRTFs were measured for twelve positions on the horizontal plane only. They were interpolated to a set of 710 positions for a better rendering of head movements and reflections. For more accurate simulations, the HRTF should be measured with a system that covers more positions, ideally with an angular resolution of 5^{o}. This would remove the need for interpolation between adjacent positions.

9.2.2. Binaural auditory system simulator

BASSIM was not able to explain the precedence effect nor was it capable to offer predictions on spaciousness. In Breebaart's model, a serie of adaptation loops were added at the end of the peripherical model. For a constant input stimulus, the adaptation reduced the level of excitation after an initial onset. This could explain forward and backward masking and some aspects of the precedence effect. In this work, adaptation was removed from the binaural model for simplicity. In future work, the influence of adaptation on the spatial prediction needs to be investigated.

Spaciousness, or environment width, is dependent on the later and diffuse reflections. It

is generally assumed that reflections arriving after 80 ms of the direct sound contribute to the perception of space. The predictions of BASSIM did not consider this attribute of the auditory system. Adding IC estimates and temporal constraints to the model could allow BASSIM to give predictions on the perceived environment width.

9.2.3. Binaural hearing aid algorithms

The algorithm presented above has been evaluated in ideal conditions. The VAD detector was assumed to be perfect and the reverberation time was known a priori. The performance of the algorithm with a more realistic voice activity and reverberation estimators need to be investigated.

Furthermore, the evaluations carried out in this thesis were based on a model of speech intelligibility and the BASSIM. Listening experiments need to be carried out in which test subjects rate the localization performance, the speech understanding and the sound quality of the algorithm. These tests have not been carried out in this project due to time constraints.

9.2.4. Perceptual evaluations

It has been argued in the introduction and in Chapter 2 that the internalization of sound is a strong limitation of current bilateral hearing prostheses. This phenomenon however has not been covered in this thesis.

Additional experiments carried out in the Phonak research laboratory investigated the perceived diffuseness and the internalization of sound sources in the same acoustical conditions as in Chapter 4. The evaluation was done in the framework of localization tests with an additional slider where the listeners had to rate these quantities. The outcome showed different results for different algorithms. This indicates that hearing aids have an impact on these aspects of spatial hearing. Further research is needed to clarify these points.

References

[Aichner *et al.* 2007] R. Aichner, H. Buchner, M. Zourub and W. Kellerman. *Multichannel source separation preserving spatial information.* In Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 15–18, 2007.

[Akeroyd *et al.* 2007] M.A. Akeroyd, S. Gatehouse and J. Blaschke. *The detection of differences in the cues to distance by elderly hearing-impaired listeners.* Journal of the acoustical society of America, vol. 121, no. 2, pages 1077–1089, February 2007.

[Akeroyd 2010] M.A. Akeroyd. *The effect of hearing-aid compression on judgments of relative distance (L).* Journal of the acoustical society of America, vol. 127, no. 1, pages 9–12, January 2010.

[ANSI-SII 1997] ANSI-SII. *American national standard methods for calculation of the speech intelligibility index ANSI S3.5-1997.* Acoustical Society of America, 1997.

[Bernstein & Trahiotis 1994] L. R. Bernstein and C. Trahiotis. *Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise.* Journal of the Acoustical Society of America, vol. 95, no. 6, pages 3561–3567, June 1994.

[Best *et al.* 2010] 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, pages 723–732, 2010.

[Beutelmann & Brand 2006] R. Beutelmann and T. Brand. *Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners.* Journal of the Acoustical Society of America, vol. 1, no. 120, pages 331–342, July 2006.

[Blauert & Lindemann 1986] J. Blauert and W. Lindemann. *Auditory spaciousness: Some further psychoacosutics analyses.* Journal of the Acoustic Society of America, vol. 80, no. 2, pages 533–542, August 1986.

[Blauert 2005] J. Blauert. Communication acoustics. Springer, 2005.

[Boymans *et al.* 2009] M. Boymans, S. T. Goverts, S. E. Kramer, J. M. Festen and W. A. Dreschler. *Candidacy for bilateral hearing aids: a retrospective multicenter study.* Journal of Speech, Language, and Hearing Research, vol. 52, pages 130–140, 2009.

[Breebaart *et al.* 2001] J. Breebaart, S. van de Paar and A. Kohlrausch. *Binaural processing model based on contralateral inhibition. I. Model structure.* Journal of the Acoustical Society of America, vol. 110, no. 2, pages 1074–1081, 2001.

[Breebaart 2001] J. Breebaart. Modeling binaural signal detection. PhD Thesis, TU Eindhoven, 2001.

[Breiman 1984] Leo Breiman. Classification and regression trees. Wadsworth International Group, Belmont, Calif, 1984.

[Breiman 2001] L. Breiman. *Random forests.* Machine Learning, vol. 45, no. 1, pages 5–32, 2001.

[Breiman 2002] L. Breiman. *Manual on setting up, using, and understanding random forests v3.1*, 2002.

[Brimjoin *et al.* 2010] W. Owen Brimjoin, David McShefferty and Michael A. Akeroyd. *Auditory and visual orienting responses in listeners with and without hearing impairment.* Journal of the Acoustical Society of America, vol. 127, no. 6, pages 3678–3688, June 2010.

[Bronkhorst & Plomp 1988] A. W. Bronkhorst and P. Plomp. *The effect of head-induced interaural time and level differences on speech intelligibility in noise.* Journal of the Acoustical Society of America, vol. 83, no. 4, pages 1508–1516, 1988.

[Bronkhorst 1995] A.W. Bronkhorst. *Localization of real and virtual sound sources.* Journal of the Acoustical Society of America, vol. 98, no. 5, pages 2542–2553, November 1995.

[Brungart *et al.* 1999] D.S. Brungart, N.I. Durlach and W.M. Rabinowitz. *Auditory localization of nearby sources. II. Localization of a broadband source.* Journal of the Acoustical Society of America, vol. 106, pages 1956–1968, 1999.

[Byrne *et al.* 1996] D. Byrne, W. Noble and B. Glauerdt. *Effects of earmold type on ability to locate sounds when wearing hearing aids.* Ear and Hearing, vol. 17, pages 218–228, 1996.

[Campbell *et al.* 2005] D. Campbell, K. Palomaeki and G. Brown. *A matlab simulation of shoebox room acoustics for use in research and teaching.* Computing and Information Systems Journal, vol. 9, no. 3, pages 48–51, October 2005.

[Carlile *et al.* 1997] S. Carlile, P. Leong and S. Hyams. *The nature and distribution of errors in sound localization by human listeners.* Hearing research, vol. 114, no. 1-2, pages 179–196, December 1997.

[Ching 2005] T.Y.C. Ching. *The evidence calls for making binaural-bimodal fittings routine.* The Hearing Journal, vol. 58, no. 11, pages 32–41, November 2005.

[Christensen *et al.* 1999] F. Christensen, H. Moeller, P. Minnaar, J. Plogsties and S. K. Olesen. *Interpolating between head-related transfer functions measured with low directional resolution.* In 107th AES convention, 1999.

[Damgrave & Lutters 2009] R.G.J Damgrave and D. Lutters. *The drift of the Xsens motion capturing suit during common movements in a working environment.* In Proceedings of the 19th CIR[Brimjoin *et al.* 2010]P design conference - competitive design, pages 338–343, March 2009.

[D.Griesinger 1992] D.Griesinger. *Measures of spatial impression and reverberance based on the physiology of human hearing.* Proceedings of the 11th Audio Engineering Society Conference, pages 114–145, 1992. Portland, Oregon, USA.

[Doclo & Moonen 2002] S. Doclo and M. Moonen. *GSVD-based optimal filtering for single and multimicrophone speech enhancement.* IEEE Transaction on Signal Processing, vol. 50, no. 9, pages 2230–2244, 2002.

[Dreschler *et al.* 2001] W. A. Dreschler, H. Verschuure, C. Ludvigsen and S. Westermann. *ICRA-Noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment.* Audiology, vol. 40, pages 148–157, 2001.

[Duda & Martens 1998] R.O. Duda and W.L. Martens. *Range dependance of a spherical head model.* Journal of the Acoustical Society of America, vol. 104, no. 4, pages 3048–3058, November 1998.

[Durlach 1963] N.I. Durlach. *Equalization and cancellation theory of binaural masking-level differences.* Journal of the Acoustical Society of America, vol. 35, pages 1206–1218, 1963.

[Falk & Chan 2008] T. H. Falk and W.-Y. Chan. *A non-intrusive quality measure of dereverberated speech.* International Workshop for Acoustic Echo and Noise Control, 2008.

[Faller & Merimaa 2004] C. Faller and J. Merimaa. *Source localization in complex listening situations: Selection of binaural cues based on interaural coherence.* Journal of the Acoustical Society of America, vol. 116, no. 5, pages 3075–3089, November 2004.

[Festen 1993] J.M. Festen. *Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering noise.* Journal of the Acoustical Society of America, vol. 94, no. 3, pages 1295–1300, September 1993.

[G. Jr. Kidd *et al.* 2005] T. L. Arbogast G. Jr. Kidd, C. R. Mason and F. J. Gallum. *The advantage of knowing where to listen.* Journal of the Acoustical Society of America, vol. 118, pages 3804–3815, 2005.

[Gardner & Martin 1994] B. Gardner and K. Martin. *HRTF Measurements of a KEMAR Dummy-Head Micophone.* MIT Media Lab Perceptual Computing - Technical Report, vol. 280, pages 1–7, May 1994.

[Gardner 1995] W. Gardner. *Efficient convolution without input-output delay.* Journal of the Audio Engineering Society, vol. 43, no. 3, pages 127–136, March 1995.

[Garofolo *et al.* 1993] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett and N. L. Dahlgren. *TIMIT Acoustic Phonetic Continuous Speech Corpus*, 1993.

[Gatehouse & Noble 2004] S. Gatehouse and W. Noble. *The Speech, Spatial, and Qualities of Hearing scale (SSQ).* International Journal of Audiology, vol. 43, pages 85–99, 2004.

[Gilkey & Anderson 1995] R.H. Gilkey and T.R. Anderson. *The accuracy of absolute localization judgments for speech stimuli.* Journal of Vestibular Research, vol. 5, no. 6, pages 487–497, 1995.

[Good & Gilkey 1996] M.D. Good and R.H. Gilkey. *Sound localization in noise: The effect of signal-to-noise ratio.* Journal of the Acoustical Society of America, vol. 99, no. 2, pages 1108–1117, February 1996.

[Goupell & Hartmann 2006] M.J. Goupell and W.M. Hartmann. *Interaural fluctuations and the detection of interaural iincoherence: Bandwidth effects*. Journal of the Acoustical Society of America, vol. 119, no. 6, pages 3971–3986, June 2006.

[Goupell & Hartmann 2007a] M.J. Goupell and W.M. Hartmann. *Interaural fluctuations and the detection of interaural incoherence II: Brief duration noises*. Journal of the Acoustical Society of America, vol. 121, no. 4, pages 2127–2136, April 2007.

[Goupell & Hartmann 2007b] M.J. Goupell and W.M. Hartmann. *Interaural fluctuations and the detection of interaural incoherence III: Narrowband experiments and binaural models*. Journal of the Acoustical Society of America, vol. 122, no. 2, pages 1029–1045, August 2007.

[Grämer *et al.* 2010] T. Grämer, M. F. Müller, S. Schimmel, A. Kegel and N. Dillier. *Dynamic virtual acoustical reproduction system for hearing prosthesis evaluation*. Proceedings of the 13th meeting of the German Society of Audiology, 2010.

[Grant 2001] K. W. Grant. *The effect of speechreading on masked detection thresholds for filtered speech*. Journal of the Acoustical Society of America, vol. 109, pages 2272–2275, 2001.

[Grantham 1984a] D. W. Grantham. *Discrimination of dynamic interaural intensity differences*. Journal of the Acoustical Society of America, vol. 76, no. 1, pages 71–76, July 1984.

[Grantham 1984b] D. W. Grantham. *Interaural intensity discrimination: Insentivity at 1000 Hz*. Journal of the Acoustical Society of America, vol. 75, no. 4, pages 1191–1194, April 1984.

[Griesinger 1998] D. Griesinger. *General overview of spatial impression, envelopment, localization, and externalization*. Proceedings of the 15th Audio Engineering Society Conference, pages 136–149, 1998. Copenhagen, Denmark.

[Hamacher *et al.* 2005] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder and U. Rass. *Signal processing in high-end hearing aids: State of the art, challenges, and future trends*. EURASIP Journal of Applied Signal Processing, vol. 18, pages 2915–2929, 2005.

[Hartmann & Wittenberg 1996] W. M. Hartmann and A. Wittenberg. *On the externalization of sound images*. Journal of the Acoustical Society of America, vol. 99, no. 6, pages 3678–3688, 1996.

[Hawley *et al.* 1999] M. L. Hawley, R. Y. Litovsky and H. S. Colburn. *Speech intelligibility and localization in a multi-source environment*. Journal of the Acoustical Society of America, vol. 105, no. 6, pages 3436–3448, June 1999.

[Hershkowitz & Durlach 1969] R. M. Hershkowitz and N. I. Durlach. *Interaural time and amplitude JNDs for a 500-Hz tone*. Journal of the Acoustical Society of America, vol. 56, pages 1464–1467, 1969.

[Hidaka *et al.* 1995] T. Hidaka, L. Beranek and T. Okano. *Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in*

concerts halls. Journal of the Acoustical Society of America, vol. 98, no. 2, pages 988–1007, August 1995.

[Holube *et al.* 1998] I. Holube, M. Kinkel and B. Kollmeier. *Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments.* Journal of the Acoustical Society of America, vol. 104, pages 2412–2425, 1998.

[Hoshuyama *et al.* 1999] O. Hoshuyama, A. Sugiyama and A. Hirano. *A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters.* IEEE Transactions on Signal Processing, vol. 47, pages 2677–2683, 1999.

[Ingard 1953] U. Ingard. *A review of the influence of meteorological conditions on sound propagation.* Journal of the Acoustical Society of America, vol. 25, pages 405–411, 1953.

[Jeffress 1948] L.A. Jeffress. *A place theory of sound localization.* Journal of Comparative and Physiological Psychology, vol. 41, pages 35–39, 1948.

[Jeub *et al.* 2010] M. Jeub, M. Schäfer, T. Each and P. Vary. *Model-based dereverberation preserving binaural cues.* IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 7, pages 1732–1745, 2010.

[Keidser *et al.* 2006] G. Keidser, K. Rohrseitz, H. Dillon, V. Hamacher, L. Carter, U. Rass and E. Convery. *The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal performance in hearing aid wearers.* International Journal of Audiology, vol. 45, pages 563–579, 2006.

[Kerber & Seeber 2009] Stefan Kerber and Bernhard U. Seeber. *Compatibility of a magnetic position tracker with a cochlear implant system.* Ear and Hearing, vol. 30, pages 380–383, 2009.

[Kim & Choi 2005] S. Kim and W. Choi. *On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach.* Journal of the acoustical Society of America, vol. 117, no. 6, pages 3657–3665, June 2005.

[Klump & Eady 1956] R.G. Klump and H.R. Eady. *Some measurements of interaural time difference thresholds.* Journal of the Acoustical Society of America, vol. 5, no. 28, pages 859–860, September 1956.

[Knapp & Carter 1976] C. H. Knapp and G. C. Carter. *The generalized correlation method for estimation of time delay.* IEEE Transaction on Acoustics, Speech and Signal Processing, vol. 24, no. 4, pages 320–327, 1976.

[Köbler & Rosenhall 2002] S. Köbler and U. Rosenhall. *Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification.* International Journal of Audiology, vol. 41, pages 395–400, 2002.

[Langendijk & Bronkhorst 2000] E.H.A. Langendijk and A.W. Bronkhorst. *Fidelity of three-dimensional-sound reproduction using a virtual auditory display.* Journal of the Acoustical Society of America, vol. 107, no. 1, pages 528–537, January 2000.

[Langendijk *et al.* 2001] E. H. A. Langendijk, D. J. Kistler and F. L. Wightman. *Sound localization in the presence of one or two distracters.* Journal of the Acoustical Society of America, vol. 109, no. 5, pages 2123–2134, May 2001.

[L.E. Kinsler & Sanders 2000] A.B. Coppens L.E. Kinsler A.R. Frey and J.V. Sanders. Fundamentals of acoustics, 4th edition. John Wiley and Sons, Inc., 2000.

[Lentz *et al.* 2007] T. Lentz, D. Schrï£¡der, M. Vorlï£¡nder and I. Assenmacher. *Virtual reality system with integrated sound field simulation and reproduction.* EURASIP Journal on Advances in Signal Processing, pages 1–19, 2007.

[Lindemann 1986a] W. Lindemann. *Extension of a binaural cross-correlation model by contralateral inhibition. I. simulation of lateralization for stationary signal.* Journal of the Acoustical Society of America, vol. 80, page 1608âĂŞ1622, 1986.

[Lindemann 1986b] W. Lindemann. *Extension of a binaural cross-correlation model by contralateral inhibition. II. the law of the ïňĄrst wavefront.* Journal of the Acoustical Society of America, vol. 80, page 1623âĂŞ1630, 1986.

[Litovsky *et al.* 1999] R. Y. Litovsky, H. S. Colburn, W. A. Yost and S. J. Guzman. *The precedence effect.* Journal of the Acoustical Society of America, vol. 106, pages 1633–1654, 1999.

[Long 2000] Christopher J. Long. *Bilateral cochlear implants: Basic psychophysics.* PhD thesis, MIT, September 2000.

[Lorenzi *et al.* 1999] C. Lorenzi, S. Gatehouse and C. Lever. *Sound localization in noise in normal-hearing listeners.* Journal of the Acoustical Society of America, vol. 105, no. 3, pages 1810–1820, March 1999.

[Mackensen 2004] P. Mackensen. *Auditive Localization. Head movements, an additional cue in localization.* PhD thesis, TU Berlin, 2004.

[Macpherson & Middlebrooks 2002] E. A. Macpherson and J.C. Middlebrooks. *Listener weighting of cues for lateral angle: The duplex theory of sound lateralization revisited.* Journal of the Acoustical Society of America, vol. 111, no. 5, pages 2219–2236, May 2002.

[Makous & Middlebrooks 1990] J.C. Makous and J.C. Middlebrooks. *Two-dimensional sound localization by human listeners.* Journal of the Acoustical Society of America, vol. 87, no. 5, pages 2186–2200, May 1990.

[Marzinik & Kollmeier 2004] M. Marzinik and B. Kollmeier. *Speech pause detection for noise spectrum estimation by tracking power evelope dynamics.* IEEE Transactions on Speech Audio Processing, vol. 10, no. 2, pages 109–118, 2004.

[Mason *et al.* 2005] R. Mason, T. Brookes and F. Rumsey. *Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli.* Journal of the Acoustical Society of America, vol. 117, no. 3, pages 1337–1350, March 2005.

[Mason 2002] R. Mason. *Elicitation and measurement of auditory spatial attributes in reproduced sounds.* PhD thesis, University of Surrey, February 2002.

[Matsumoto *et al.* 2004a] M. Matsumoto, S. Yamanaka, M. Tohyama and H. Nomura. *Effect of arrival time correction on the accuracy of binaural impulse response interpolation. Interpolation methods of binaural response.* Journal of the Audio Engineering Society, vol. 52, no. 1/2, pages 56–61, January-February 2004.

[Matsumoto *et al.* 2004b] M. Matsumoto, S. Yamanaka, M. Tohyama and H. Nomura. *Effect of arrival time correction on the accuracy of binaural room impulse response interpolation. Interpolation of room impulse responses.* Journal of the Audio Engineering Society, vol. 52, pages 56–61, 2004.

[Mershon & King 1975] D.H. Mershon and E. King. *Intensity and reverberance as factors in the auditory perception of egocentric distance.* Perception Psychophysics, vol. 18, pages 409–415, 1975.

[Middlebrooks & Green 1990] J. C. Middlebrooks and D. M. Green. *Directional dependance of interaural envelope delays.* Journal of the Acoustical Society of America, vol. 87, no. 5, pages 2149–2162, May 1990.

[Moeller 1992] H. Moeller. *Fundamentals of binaural technology.* Applied Acoustics, vol. 36, pages 171–218, 1992.

[Müller *et al.* 2010] M. F. Müller, A. Kegel, S. Schimmel and N. Dillier. *Localization of virtual sound sources in realistic and complex scenes: how much do hearing aids alter localization.* Proceedings of the 13th meeting of the German Society of Audiology, 2010.

[Müller *et al.* 2011a] M. F. Müller, T. Grämer, S. Schimmel and N. Dillier. *Efficient reproduction of head movements and dynamic scenes using virtual acoutics.* accepted for publication in Applied Acoustics, 2011.

[Müller *et al.* 2011b] M. F. Müller, A. Kegel, M. Hofbauer, S. Schimmel and N. Dillier. *Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes.* accepted for publication in Journal of the Acoustical Society of America, 2011.

[Müller *et al.* 2011c] M. F. Müller, K. Meisenbacher, W.-K. Lai, and N. Dillier. *Influence of head movements on sound localization with cochlear implants.* Proceedings of the 14th meeting of the German Society of Audiology, 2011.

[Müller *et al.* 2011d] M. F. Müller, K. Meisenbacher, W.-K. Lai, and N. Dillier. *Sound localization with bilateral cochlear implants in noise; how much do head movements help for localization?* Submitted to Cochlear Implants International, 2011.

[Munhall *et al.* 2004] K.G. Munhall, Jeffery A. Jones, Daniel E. Callan, Takaaki Kuarate and Eric Vatikiotis-Bateson. *Visual prosody and speech intelligibility: Head movement improves auditory speech perception.* Psychological Science, vol. 15, pages 133–137, 2004.

[Noble & Byrne 1990] W. Noble and D. Byrne. *A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical plane.* British Journal of Audiology, vol. 24, pages 335–346, 1990.

[Noble & Gatehouse 2006] W. Noble and S. Gatehouse. *Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing scale (SSQ).* International Journal of Audiology, vol. 45, pages 172–181, March 2006.

[Nuetzel & Hafter 1981] J. M. Nuetzel and E. R. Hafter. *Discrimination of interaural delays in complex waveforms: Spectral effects.* Journal of the Acoustical Society of America, vol. 69, no. 4, pages 1113–1118, April 1981.

[Ramirez *et al.* 2007] J. Ramirez, J. M. Goeritz and J. C. Segura. Voice activity detection. fundamentals and speech recognition system robustness, volume 62. 2007.

[Ramirez 2004] J. Ramirez. *Efficient voice activity detection algorithms using longterm speech information.* Speech Communication, vol. 42, pages 271–287, 2004.

[Raspaud *et al.* 2010] M. Raspaud, H. Viste and G. Evangelista. *Binaural source localization by joint estimation of ILD and ITD.* IEEE Transaction on Audio, Speech and Language Processing, vol. 18, no. 1, pages 68–77, 2010.

[Reed & Blum 1990] M.C. Reed and J.J. Blum. *A model for the computation and encoding of azimuthal information by the lateral superior olive.* Journal of the Acoustical Society of America, vol. 88, no. 3, pages 1442–1453, 1990.

[Reindl *et al.* 2010] K. Reindl, Y. Zheng and W. Kellermann. *Speech enhancement for binaural hearing aids based on blind source separation.* In Proc. of 4th International Symposium on Communication, Control and Signal Processing (ISCCSP), pages 241–244, 2010.

[Rife & Vanderkooy 1989] D.D. Rife and J. Vanderkooy. *Transfer-Function Measurements with Maximum-Length Sequences.* Journal of the Audio Engineering Society, vol. 37, no. 6, pages 419–444, June 1989.

[Rychtarikova *et al.* 2009a] M. Rychtarikova, T. van den Bogaert, G. Vermeir and J. Wouters. *Binaural sound source localization in real and virtual rooms.* Journal of the Audio Engineering Society, vol. 57, no. 4, pages 205–220, April 2009.

[Rychtarikova *et al.* 2009b] M. Rychtarikova, T. van den Bogaert, G. Vermeir and J. Wouters. *Binaural sound source localization in real and virtual rooms.* Journal of the Audio Engineering Society, vol. 57, pages 205–220, 2009.

[Saberi 1998] K. Saberi. *Modeling interaural-delay sensitivity to frequency modulation a high frequencies.* Journal of the Acoustical Society of America, vol. 103, no. 5, pages 2551–2564, May 1998.

[Sakamoto *et al.* 1976] N. Sakamoto, T. Gotoh and Y. Kimura. *On out.of.head localization in headphone listening.* Journal of the Audio Engineering Society, vol. 24, pages 710–716, 1976.

[Savioja *et al.* 1999] L. Savioja, J. Huopaniemi, T. Lokki and R. Vï£¡ï£¡nï£¡nen. *Creating interactive virtual acoustic environments.* Journal of the Audio Engineering Society, vol. 47, no. 9, pages 675–705, September 1999.

[Schimmel *et al.* 2009] S.M. Schimmel, M.F. Müller and N. Dillier. *A fast and accurate "shoe-box" room acoustics simulator.* In Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 241–244, 2009. http://roomsim.sourceforge.net.

[Seeber & Fastl 2008] B. U. Seeber and H. Fastl. *Localization cues with cochlear implants.* Journal of the acoustical society of america, vol. 123, no. 2, pages 1030–1042, February 2008.

[S.Perrett & Noble 1997] S.Perrett and W. Noble. *The contribution of head motion cues to localization of low-pass noise.* Perception and Psychophysics, vol. 59, no. 7, pages 1018–1026, 1997.

[Spriet *et al.* 2004] A. Spriet, M. Moonen and J. Wouters. *spatially pre-processed speech distortion weighted multi-channel Wiener.* IEEE Transaction on Signal Processing, vol. 84, no. 12, pages 2367–2387, 2004.

[Stern & Colburn 1978] R. M. Stern and H. S. Colburn. *Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position.* Journal of the Acoustical Society of America, vol. 64, no. 1, pages 127–140, 1978.

[Stern & Trahiotis 2001] R. M. Stern and C. Trahiotis. *Manipulating the straghtness and curvature of patterns of interaural cross correlation affects listeners'sensitivity to changes in interaural delay.* Journal of the Acoustical Society of America, vol. 109, no. 1, pages 321–330, January 2001.

[Stern *et al.* 1988] R. M. Stern, A. S. Zeiberg and C. Trahiotis. *Lateralization of complex binaural stimuli: A weighted image model.* Journal of the Acoustical Society of America, vol. 84, no. 1, pages 156–165, July 1988.

[Technologies 2010] XSens Technologies. *http://www.xsens.com*, 2010.

[Toole 1969] F. E. Toole. *In-head localization of acoustic images.* Journal of the Acoustical Society of America, vol. 48, pages 943–949, 1969.

[Trahiotis & Stern 1988] C. Trahiotis and R. M. Stern. *Across-frequency interaction in lateralization of complex binaural stimuli.* Journal of the Acoustical Society of America, vol. 84, no. 1, pages 156–165, July 1988.

[Tschopp & Ingold 1992] K. Tschopp and L. Ingold. Moderne verfahren der sprachaudiometrie, chapitre Die Entwicklung einer deutschen Version des SPIN-Tests (speech perception in noise), pages 311–329. Median-Verlag von Killisch-Horn,Heidelber, 1992.

[Tucci *et al.* 2010] D. L. Tucci, M. H. Merson and B. Wilson. *A Summary of the Literature on Global Hearing Impairment: Current Status and Priorities for Action.* Otology & Neurotology, vol. 31, no. 1, pages 31–41, 2010.

[van den Bogaert *et al.* 2006] T. van den Bogaert, T.J. Klasen, M. Moonen, L. Van Deun and J. Wouters. *Horizontal localization with bilateral hearing aids: Without is better than with.* Journal of the Acoustical Society of America, vol. 116, no. 1, pages 515–526, 2006.

[van den Bogaert *et al.* 2011] 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, pages 164–176, 2011.

[van den Bogaert 2008] T. van den Bogaert. *Preserving binaural cues in noise reduction algorithms for hearing aids.* PhD thesis, KUL Leuven, 2008.

[van Hoesel & Tyler 2003] Richard J. M. van Hoesel and Richard S. Tyler. *Speech perception, localization, and lateralization with bilateral cochlear implants.* Journal of the acoustical Society of America, vol. 113, no. 3, pages 1617–1630, March 2003.

[van Hoesel 2004] Richard J. M. van Hoesel. *Exploring the benefits of bilateral cochlear implants.* Audiology and neuro-otology, vol. 9, pages 234–246, 2004.

[Wagener & Brand 2005] K.C. Wagener and T. Brand. *Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters.* International Journal of Audiology, vol. 44, no. 3, pages 144–156, 2005.

[Wallach 1940] H. Wallach. *The role of head movements and vestibular and visual cues in sound localization.* Journal of Experimental Psychology, vol. 27, no. 4, pages 339–368, October 1940.

[Wang & Brown 2006] D. Wang and G.J. Brown. Computational auditory scene analysis: Principles, algorithms and applications. Wiley Interscience, 2006.

[Wenzel *et al.* 1993] E. M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wightman. *Localization using nonindividualized head-related transfer functions.* Journal of the Acoustical Society of America, vol. 94, no. 1, pages 111–123, July 1993.

[Wightman & Kistler 1989] F.L. Wightman and D.J. Kistler. *Headphone simulation of free-field listening. II: Psychophysical validation.* Journal of the Acoustical Society of America, vol. 85, no. 2, pages 868–878, February 1989.

[Wightman & Kistler 1999] F.L. Wightman and D.J. Kistler. *Resolution of front-back ambiguity in spatial hearing by listener and source movement.* Journal of the Acoustical Society of America, vol. 105, no. 5, pages 2841–2853, May 1999.

[Wittkop & Hohmann 2003] T. Wittkop and V. Hohmann. *Strategy selective noise reduction for binaural digital hearing aids.* Speech Communication, vol. 39, no. 2, pages 111–138, 2003.

[Wittkop 2001] T. Wittkop. Two-channel noise reduction algorithms motivated by models of binaural interaction. University Oldenburg, PhD Thesis, 2001.

[Yin 2002] T.C.T Yin. Integrative functions in the mammalian auditory pathway, chapitre 4-Neural mechanisms of encoding binaural localization cues in the auditory brainstem, pages 99–159. Springer handbook of auditory research, 2002.

[Yost & Dye 1988] W. A. Yost and R.H. Dye. *Discrimination of interaural differences of level as function of frequency.* Journal of the Acoustical Society of America, vol. 83, no. 5, pages 1846–1851, May 1988.

[Zahorik 2002] P. Zahorik. *Assessing auditory distance perception using virtual acoustics.* Journal of the Acoustical Society of America, vol. 111, no. 4, pages 1832 – 1846, April 2002.

[Zwicker & H.Fastl 1990] E. Zwicker and H.Fastl. Psychoacoustics, facts and models. Springer-Verlag, 1990.

[Zwislocki & Feldman 1956] J. Zwislocki and R. S. Feldman. *Just Noticeable Differences in Dichotic Phase*. Journal of the Acoustical Society of America, vol. 5, no. 28, pages 860–864, September 1956.

Curriculum Vitae

Martin Felix Müller was born in Lausanne, Switzerland, in 1982. In 2006 he received the Master of Science degree in Communication Systems from the Swiss Federal Institute of Technology in Lausanne (EPFL). During his studies, he stayed one year at the Chalmers University of Technology, Göteborg, Sweden, in 2004. In 2006, he spent six months doing research for his Master Thesis at Philips Research Laboratories in Eindhoven, the Netherlands.Since 2007 he has been working towards a PhD degree in Science at the Laboratory of Experimental Audiology of the University Hospital Zurich and the Institute for Biomedical Engineering of the Swiss Federal Institute of Technology Zurich (ETHZ), Switzerland, on a joint project with the hearing instrument manufacturer Sonova (former Phonak). His main research interests include audio signal processing, human auditory perception and rehabilitive audiology.

List of Publications

JOURNAL PAPERS

1. M.F. Müller, A. Kegel, M. Hofbauer, S. Schimmel and N. Dillier, "Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes", in *Journal of the Acoustical Society of America*, 2011, accepted for publication.
2. M.F. Müller, T. Grämer, S. Schimmel and N. Dillier, "Efficient reproduction of head movements and dynamic scenes using virtual acoutics", in *Applied Acoustics*, 2011, submitted.
3. M.F. Müller, K. Meisenbacher, W.-K. Lai and and N. Dillier, "Sound localization with bilateral cochlear implants in noise; how much do head movements help for localization?", in *Cochlear Implants International*, 2011, submitted.

CONFERENCE PAPERS

1. S. Schimmel, M.F. Müller and N. Dillier, "A fast and accurate "shoebox" room acoustics simulator", in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.
2. M.F. Müller, A. Kegel, S. Schimmel and N. Dillier, "Localization of virtual sound sources in realistic and complex scenes: how much do hearing aids alter localization", in *Proceedings of the 13th meeting of the German Society of Audiology*, 2010.
3. T. Grämer, M.F. Müller, A. Kegel, S. Schimmel and N. Dillier, "Dynamic virtual acoustical reproduction system for hearing prosthesis evaluation", in *Proceedings of the 13th meeting of the German Society of Audiology*, 2010.
4. M.F. Müller, Meisenbacher, W.-K. Lai and and N. Dillier, "Influence of head movements on sound localization with cochlear implants", in *Proceedings of the 14th meeting of the German Society of Audiology*, 2011.