Individualized directional microphone optimization in hearing aids based on reconstructing the 3D geometry of the head and ear from 2D images

Harder, Stine

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Individualized directional microphone optimization in hearing aids based on reconstructing the 3D geometry of the head and ear from 2D images

Stine Harder

Kongens Lyngby 2015
PhD-2015-365

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Matematiktorvet, building 303B,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3351
compute@compute.dtu.dk
www.compute.dtu.dk PhD-2015-365

Summary (English)

The goal of this thesis is to improve intelligibility for hearing-aid users by individualizing the directional microphone in a hearing aid. The general idea is a three step pipeline for easy acquisition of individually optimized directional filters. The first step is to estimate an individual 3D head model based on 2D images, the second step is to simulate individual head related transfer functions (HRTFs) based on the estimated 3D head model and the final step is to calculate optimal directional filters based on the simulated HRTFs. The pipeline is employed on a Behind-The-Ear (BTE) hearing aid.

We verify the directional filters optimized from simulated HRTFs based on a listener-specific head model against two set of optimal filters. The first set of optimal filters is calculated from HRTFs measured on a 3D printed version of the head model. The second set of optimal filters is calculated from HRTFs measured on the actual human subject.

A verification of the 'simulated' directional filters against the optimal filters for the human subject revealed a 0.5 dB reduction in articulation-index weighted directivity index, which corresponds to 5% less speech intelligibility. A comparison against non-individual directional filters revealed equally high Articulation-Index weighted Directivity Index (AI-DI) values for our specific test subject. However, measurements on other individuals indicate that the performance of the non-individual filters vary among subjects, and in particular individuals who deviate from an average of the population could benefit from having individualized filters.

We developed a pipeline for 3D printing of full size human heads. The 3D printed

head facilitated the second verification step, which revealed a 0.3 dB reduction from optimal to simulated directional filters. This indicates that the simulation are more similar to measurements on the 3D printed head than measurements on the human subject. We suggest that the larger difference between simulation and human measurements could arise due to small geometrical errors in the head model or due to differences in acoustical properties between human skin and virtual material properties in the simulation.

The BTE hearing aid showed very little room for improvement using individualized directional filters, however the directional filters in an In-The-Ear (ITE) hearing aid revealed an improvement in AI-DI values of up to 3.6 dB between an average filter and an optimal filter. This suggests that hearing-aid users with ITE hearing aids could benefit more from having individualized directional filters than what was shown for a BTE hearing aid.

This thesis is a step towards individualizing the directional microphone in hearing aids, which could contribute with improved sound for a group of hearing-aid users. In particular, we believe that ITE hearing-aid users could have a large benefit from an individualized directional microphone.

Summary (Danish)

Målet med denne afhandling er at forbedre høreapparatsbrugeres taleforståelighed ved at individualisere retningsmikrofonen i et høreapparat. Den overordnede ide er en tretrins pipeline, der muliggør en nemmere tilpasning af individuelt optimerede retningsvirknings-filtre. Det første trin er at estimere en individuel 3D hovedmodel ud fra 2D billeder, det andet trin er at simulere individuelle head related transfer functions (HRTFs) baseret på den estimerede 3D hovedmodel og det sidste trin er at beregne optimale filtre med retningsvirkning baseret på de simulerede HRTFs. Pipelinen testes på et Behind-The-Ear (BTE) høreapparat.

Vi verificerer retningsvirknings-filtrene, der er optimeret på HRTFs simuleret ud fra en individuel hovedmodel, på to sæt af optimale filtre. Det første sæt af optimale filtre er beregnet ud fra HRTFs målt på en 3D printet version af hovedmodellen. Det andet sæt optimale filtre er beregnet ud fra HRTFs målt på den faktiske testperson.

Verifikationen af de 'simulerede' retningsvirknings-filtre imod de optimale filtre, beregnet ud fra HRTFs målt på testpersonen, viste en reduktion på 0.5 dB i Articulation-Index vægtet Directivity Index (AI-DI), hvilket svarer til 5% mindre taleforståelighed. En evaluering af ikke-individualiserede retningsvirkningsfiltre afslørede sammenlignelige AI-DI værdier for vores specifikke testperson. Målinger på andre testpersoner viste imidlertid, at virkningen af ikke individualiserede filtre varierer blandt testindivider. Individer som afviger fra befolkningsgennemsnittet kunne i særlig grad have fordel af individualiserede retningsvirknings filtre.

Vi har udviklet en pipeline for 3D printning af menneskehovedmodeller i fuld størrelse. Det andet verifikationstrin blev muliggjort af vores 3D printede hovedmodel og det afslørede en reduktion på 0.3 dB fra optimale til simulerede retningsvirknings-filtre. Dette indikerer, at simuleringen ligner målingerne på det printede hoved i højere grad end målingerne på testpersonen. Vi formoder, at de større forskelle imellem simuleringer og målinger på testpersonen skyldes små geometriske fejl i hovedmodellen eller forskelle i de akustiske egenskaber i mellem hud og de virtuelle materialer i simuleringen.

For BTE høreapparartet blev der konstateret en meget lille mulighed for forbedring ved at bruge individualiserede retningsvirknings-filtre. Retningsvirkningsfiltrene i et In-The-Ear (ITE) høreapparart viste en forbedring på op til 3.6 dB fra et gennemsnits filter til et optimalt filter. Dette tyder på, at høreapparatsbrugere, der benytter et ITE høreapparat, kunne opnå en større fordel ved at have individualiserede filtre end hvad der blev konstateret for et BTE høreapparat.

Denne afhandling er et skridt i retning mod en individualisering af retningsmikrofonen i høreapparater, hvilket kunne medføre forbedret lyd for en gruppe af høreapparatsbrugere. Vi forventer at specielt høreapparats brugere med et ITE høreapparat vil have en stor fordel af individualliserede retningsmikrofoner.

Preface

This thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfilment of the requirements for acquiring a PhD in Applied Mathematics. The thesis was carried out with funding solely from the Oticon Foundation with associate professor Rasmus R. Paulsen as main supervisor and Søren Laugesen from The Eriksholm Research Centre as co-supervisor.

The thesis deals with individual optimization of the directional microphone in hearing aids based on the individual's head and ear shape.

The thesis comprise a theoretical and methodological explanation of the work that has been carried out during the PhD project. The thesis includes a total of four papers.

Lyngby, 12-March-2015

Stine Harder

Contributions

Papers included in the thesis

[A] Stine Harder, Rasmus R Paulsen, Martin Larsen, Søren Laugesen A three dimensional children head database for acoustical research and development, *Meetings on Acoustics. Proceedings (ISBN: 1939-800X) vol:19, 2013.*

[B] Stine Harder, Rasmus R Paulsen, Martin Larsen, Søren Laugesen, Michael Mihocic, Piotr Majdak A framework for comparison between measured and simulated head-related transfer functions. *Submitted to Computer-Aided Design (Journal).*

[C] Stine Harder, Rasmus R Paulsen, Martin Larsen, Søren Laugesen, Michael Mihocic, Piotr Majdak Reliability in measuring head related transfer functions of hearing aids. *Submitted to ACTA Acoustica united with Acoustica (Journal).*

[D] Stine Harder, Rasmus R Paulsen, Martin Larsen, Michael S Pedersen, Søren Laugesen, Michael Mihocic, Piotr Majdak Optimizing hearing-aid directionality from measurements and simulations. *In submission: The paper will be submitted to The Journal of the Acoustical Society of America after final remarks from the co-authors.*

Other papers by the author

- C. Pietroni, J. D. Andersen, P. Johansen, S. Harder, R. R. Paulsen, C. Børsting, N. Morling. The genetics of eye colours in an Italian population measured with an objective method for eye colour quantification, *Forensic Science International: Genetics. (ISSN: 1875-1768) vol: 4, issue: 1, pages e23-e24, 2013.*

- Stine Harder, Line K. H. Clemmensen, Anders B. Dahl, Jeppe D. Andersen, Peter Johansen, Susanne R. Christoffersen, Niels Morling, Claus Børsting, Rasmus R. Paulsen. Correlation of iris biometrics and DNA, *In proceedings of International Workshop on Biometrics and Forensics (ISBN: 978-1-4673-4987-1) IEEE, 2013.*

- Jeppe D. Andersen, Peter Johansen, Stine Harder, Susanne R. Christoffersen, Mikaela C. Delgado, Sarah T. Henriksen, Mette M. Nielsen, Erik Sørensen, Henrik Ullum, Thomas Hansen, Anders B. Dahl, Rasmus R. Paulsen, Claus Børsting, Niels Morling. Genetic analyses of the human eye colours using a novel objective method for eye colour classification, *Forensic Science International: Genetics. (ISSN: 1872-4973) vol: 7, issue: 5, pages 508-515, 2013.*

- Stine Harder, Susanne R. Christoffersen, Peter Johansen, Claus Børsting, Niels Morling, Jeppe D. Andersen, Anders B. Dahl, Rasmus R. Paulsen. What Genes Tell about Iris Appearance, *In proceedings of 15th International Conference on Medical Image Computing and Computer Assisted Intervention, 2013.*

- Carlotta Pietroni, Jeppe D. Andersen, Peter Johansen, Mikkel M. Andersen, Stine Harder, Rasmus R. Paulsen, Claus Børsting, Niels Morling. The effect of gender on eye colour variation in European populations and an evaluation of the IrisPlex prediction model, *Forensic Science International: Genetics. (ISSN: 1872-4973) vol: 11, pages 1-6, 2014.*

- Tomi Huttunen, Antti Vanne, Stine Harder, Rasmus R. Paulsen, Sam King, Lee Perry-Smith, Leo Karkkainen. Rapid generation of personalized HRTFs, *In proceedings of 55th International Conference on Spatial Audio 2014.*

- Jens Fagertun, Stine Harder, Anders Rosengren, Christian Moeller, Thomas Werge, Rasmus R Paulsen, Thomas F Hansen. 3D facial landmarks: Inter-operator variability of manual annotation, *BMC Medical Imaging (ISSN: 1471-2342) vol: 14, issue: Art.No. 35, 2014.*

Acknowledgements

A very special thanks goes to my supervisors Rasmus R. Paulsen and Søren Laugesen for their great commitment to my work and for many hours of supervision. I appreciate all of our discussions and your never ending guidance. I would also like to thank Martin Larsen from Oticon A/S for his great enthusiasm and involvement in my thesis work. You have been like a third supervisor to me.

I would like to thank all of my colleagues at the Image Analysis and Computer Graphics Group at DTU and at the Eriksholm Research Centre for providing an enjoyable and collaborative working environment, which I am pleased to stay a part of for an additional two years. Jens Fagertun deserves a thanks for helping me in the early days of my project.

The Oticon foundation deserves a great thanks for funding this project.

I would also like to thank the members of the Acoustics Research Institute (ARI) in Austria for their great hospitality during our visit in Vienna. Piotr Majdak and Michael Mihocic deserves a special thanks for their great efforts during our prolonged measurement sessions and for their dedication with our subsequent collaboration.

The people at the Vision Group at University of California, San Diego (UCSD) deserves my appreciation for taking good care of me during my external stay. In particular, I would like to thank professor Serge J. Belongie for hosting and supervising me. The research stay was a great experience both professionally and personally. I would also like to thank Søren, Sanne, Rasmus, Maria, and

Alfred for making San Diego feel like home.

I would also like to thank Thomas Hansen for making the 3D surface scanner available, and also Thomas Werge, the Danish National Advanced Technology foundation and the Lundbeck Foundation for funding the scanner. Furthermore, the staff at Enheden for Psykiatrisk Forskning at Aalborg Psykiatriske Sygehus, and at Distriktspsykiatrisk Center Østerbro deserves a thanks for their hospitality in connection with our scan sessions.

A thanks goes to Marie Elgaard Korsholm Nielsen for her help with creating the 'DTU Compute 3D head database'.

Many people at Oticon A/S deserve my recognition, among these are Michael Syskind Pedersen, Johannes Steininger, Jonas Lerche Olesen, Peter Frederiksen, Karsten Bo Rasmussen and Thomas Behrens. I appreciate all of your contributions to my work.

I would also like to express my gratitude to G.R.A.S. Sound and Vibration for providing me with the geometry of KEMAR. The geometry has been very useful.

My appreciation also goes to those of my colleagues and good friends who have contributed in preparing this thesis, including Jacob Schack Vestergaard, Susanne Rytter Christoffersen, Line Clemmensen, Jannik Boll Nielsen, Hans Martin Kjer, and Morten Nobel-Jørgensen. Lenabot and all of her followers also deserve a warm thought for keeping my caffeine level high.

Finally, a loving thanks goes to my family for providing me with 25 hours a day, and to my son Oliver for 'doing homework' with his mom during busy times and for being my favorite test subject.

Contents

List of abbreviations and acronyms

2D	Two-dimensional
3D	Three-dimensional
3DMM	3D Morphable Model
AI	Articulation Index
AI-DI	Articulation-Index weighted Directivity Index
AM	Additive Manufacturing
ARI	Acoustics Research Institute
AS-DI	'ASymptotical' Directivity Index
BEM	Boundary Element Method
BTE	Behind-The-Ear
CAD	Computer-Aided Design
CIC	Completely-In-Canal
CT	Computed Tomography
DI	Directivity Index
DTU	The Technical University of Denmark

ETF	Equipment transfer function
F.O.	Front-omni
FEM	Finite Element Method
H.C.	Hyper-cardioid
HATS	Head And Torso Simulator
HRTF	Head Related Transfer Function
ICP	Iterative Closest Point
H.A.	Human Average
IID	Interaural Intensity Differences
ITC	In-The-Canal
ITD	interaural Time Difference
ITE	In-The-Ear
KB	Kongebakken
L	Left
LF	Left Front
LR	Left Rear
MESM	Multiple Exponential Sweep Method
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
NURBS	Non-Uniform Rational B-Splines
Opt.	Optimal
PC	Principal Component
PCA	Principal Component Analysis
PCL	Point Cloud Library
PFH	Point Feature Histogram
PML	Perfectly Matched Layer
R	Right

RANSAC	RANdom SAmple Consensus
RF	Right Front
RITE	Receiver-In-The-Ear
RR	Right Rear
SD	Standard Deviation
Sim.	Simulated
SLS	Selective Laser Sintering
STL	STereoLithography (file format)
TPS	Thin Plate Spline
UI	Unidirectional Index

Chapter 1

Introduction

The human hearing is an important faculty that has many vital functions, such as speech communication and notification functions. Speech communication is important for an individual's social life, and hearing-impaired individuals often feel more isolated than individuals who have lost their sense of smell or faculty of sight [Bjå04]. Notification sounds, such as the doorbell, the phone, and warning sounds in the traffic, are also highly important in everyday life. A hearing-impaired individual has difficulties with hearing all of these important sounds, which often leads to a loss in quality of life.

A hearing aid can, to some extent, rectify a hearing impairment and thereby improve the quality of life for hearing-impaired individuals. One could think that the function of a hearing aid is to simply amplify sounds arriving at the ear canal. However, most hearing impairments require more than a simple amplification of sound. Most hearing-impaired individuals have a reduced sensitivity to sounds, however they often also have a reduced dynamic range, a decreased frequency resolution, a decreased temporal resolution and a reduced ability to benefit from binaural hearing[1]. These five areas are referred to as 'the five dimensions of hearing' and they all lead to a reduced speech intelligibility [EW05].

A hearing aid is therefore constructed to rectify the five dimensions of hearing in the best possible manner, so that the intelligibility is increased. A hearing

[1] Binaural hearing means hearing with two ears.

aid contains an amplifier that alleviates the reduced sensitivity and a compressor that addresses the reduced dynamic range. The compressor can, to some extent, also correct for a decreased temporal resolution. However, these basic functions of a hearing aid are not able to overcome all of the problems that a hearing impairment introduces. Commercial hearing aids are therefore equipped with 'helping systems', whose function are to improve the signal-to-noise ratio for the sounds that enter the hearing aids. One of these helping systems is the directional microphone. A directional microphone works by suppressing sounds arriving from some directions, while maintaining sounds arriving from a specific direction, often the front. This is an advantage, since it can often be assumed that the target sound arrives from the front, because the hearing-aid user would look in the direction of, for instance, the person who is speaking. Noise on the other hand is assumed to arrive uniformly from all directions. The importance of the directional microphone is stated by Dillon [Dil12], who writes that directional microphones is the only form of signal processing that can improve the signal-to-noise ratio in a way that leads to improved intelligibility.

The directional microphone in a hearing aid is often a first-order subtractive directional microphone, where the output from two microphones are combined. This thesis has a focus on Behind-The-Ear (BTE) and In-The-Ear (ITE) hearing aids, which are a common hearing-aid models. Both the BTE and ITE contain a front and a rear microphone, and the directional microphone is obtained by subtracting a time delayed version of the rear microphone's output from the front microphone. The amount of delay on the rear microphone determines the sensitivity pattern of the directional microphone. The delay is implemented using a set of digital filters, from now on denoted 'directional filters'. The performance of a directional microphone is usually measured with the Directivity Index (DI), which describes the sensitivity towards sounds arriving from the front relative to an average of sounds from all directions [Dil12]. The frequency dependent DI is often weighted across frequency bands according to the importance of each frequency band with respect to speech, using the Articulation index (AI). The AI weighted DI then becomes the Articulation-Index weighted Directivity Index (AI-DI).

The directional microphone is, in modern hearing aids, realized using directional filters designed once for each hearing-aid model. The goal is to achieve the highest AI-DI average across the population. This can be obtained by optimizing the directional filters with respect to AI-DI using Head Related Transfer Functions (HRTFs) measured on a head and torso simulator, such as the Brüel & Kjær head and torso simulator (HATS) or KEMAR by G.R.A.S. However, a directional microphone based on average directional filters might not perform very well for all individuals. Valente et al. [VFP95] reported a large variation in individual benefit from directional microphones, which suggests that non-individual directional filters might be less than optimal for the individual hearing-aid user.

The solution to obtaining optimal directionality for the individual hearing-aid user is, in principle, straightforward. Acoustical measurements taken from a large number of loudspeakers positioned on a sphere around the hearing-aid user can be used for individual optimization of the directional filters [Ras07, LRC03]. Such measurements are, however, very cumbersome and completely intractable in clinical practice.

It has been proposed by Rasmussen [Ras07] to replace the aforementioned elaborate acoustical measurements with numerical simulations, for instance based on the Finite Element Method (FEM). This approach requires an accurate 3D geometrical representation of the individual hearing-aid user's head. This is, again in principle, relatively straightforward by scanning the human head using a state-of-the-art surface scanner [PL10]. It is however intractable to acquire an expensive 3D surface scanner in a typical clinic.

The possibility of obtaining 3D head models from simple 2D images is therefore attractive. Estimation of 3D structures based on 2D images is an on-going research theme in the field of multiple-view geometry [HZ03] where highly calibrated camera setups are normally used. Alternatively, strong statistical priors can be used to predict 3D structures with known statistical shape priors from one or a few 2D projections. In particular, it has proven possible to estimate the 3D anatomy of human faces from frontal photos [BV03]. This approach can potentially be extended to predict the entire head shape.

1.1 Original idea

The idea behind this thesis is to improve intelligibility for hearing-aid users by individualizing the directional microphone. We envisage a three step pipeline for easy acquisition of individually optimized directional filters, see Figure 1.1. The first step is to estimate an individual 3D head model based on 2D images, the second step is to simulate individual HRTFs based on the estimated 3D head model and the final step is to calculate optimal directional filters based on the simulated HRTFs.

Our long-term vision is that the audiologist uses his/her mobile phone or digital camera to acquire images of the customer's head and ears, at the same time as selecting the hearing aid. These images will be sent together with the order forms to the hearing-aid manufacturer, where a 3D model of the customer's head and ears will be estimated. Based on this model, individually optimized directional filters will be computed from FEM simulated HRTFs. The individual directional filters will be stored in the hearing aid, which is then returned to the

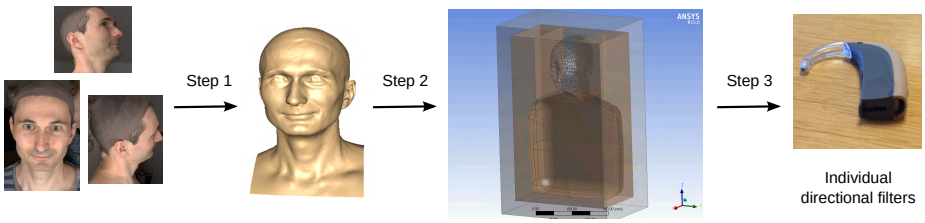


Figure 1.1: Three step project idea. Step 1: estimation of 3D model based on 2D images, step 2: FEM simulations using 3D model, and step 3: calculate optimal directional filters from simulated HRTFs.

audiologist.

1.2 Thesis outcome

Our original idea and the actual outcome of this thesis does not completely line up. The thesis has a focus on the last two parts of the originally proposed three-step pipeline. The reason being that we pursued opportunities to collaborate with interesting research partners and to investigate the areas of the project which seemed to have most novelty value and that we and our collaborators found most interesting. Chapter 5 contains the preliminary work that we performed on the topic of generating listener-specific 3D head models from 2D images.

The project was originally thought to be a collaborative work between the Eriksholm Research Center (Søren Laugesen) and the Technical University of Denmark (Rasmus Reinhold Paulsen). However, early in the project people at Oticon A/S, in particular Martin Larsen, showed interest in our work with generating 3D head models. Oticon A/S had a particular interest in children's heads, which lead to the work described in Paper A: "The DTU 3D children head database for acoustical research and development". Our 3D head models also evoked the interest of Stavros-Konstantinos Stavrakos from GN Resound A/S, with whom we collaborated on building the 'The DTU Compute 3D head database', see Chapter 4. The LocaPhoto workshop in 2012 founded a collaboration with Piotr Majdak from the Acoustics Research Institute (ARI) in Austria. This collaboration facilitated acquisition of high quality acoustical measurements in 3D, using the state-of-the-art measuring equipment at ARI. The measuring equipment at ARI was a great improvement to our own equipment, which was limited to horizontal 2D measurements. At this point 3D head models were available and in collaboration with Martin Larsen, we pur-

sued FEM simulations of HRTFs. The high quality measured HRTFs formed an ideal baseline for evaluation of the simulated HRTFs. Our final research area and goal of this thesis, was optimization of the directional microphone based on a set of HRTFs. Here we benefitted from Michael Syskind Pedersen's (Oticon A/S) experience with the topic.

During the project it became clear that the rapid development within additive manufacturing had opened up the possibility of 3D printing a real size head model for acoustical measurements. The 3D-printed head model opened up for an intermediate verification step. Besides being able to validate the directivity of a directional microphone obtained from simulated HRTFs against the optimal directivity, we were now also able to validate against the optimal directivity calculated from HRTFs measured on the 3D printed head model, see Figure 1.2. This intermediate verification step was used to exclude some of the differences that exist between HRTFs measured on a human and HRTFs simulated on a model. Verification using measurements on a printed head has the benefits that:

- Geometrical differences are limited because the simulated and measured HRTFs are based on the same head model.
- Geometrical torso differences are limited since the torso simulator by Brüel & Kjær is used for both simulation and measurements.
- Human acoustical properties from skin, hair, and cloth are replaced with plastic material that can be assumed to be more similar to the properties of the simulated model.
- Human movements, caused by for instance breathing, are removed.

Figure 1.2 summarizes the goal of this thesis, which is to individually optimize the directional microphone from simulations and it shows our two verification steps.

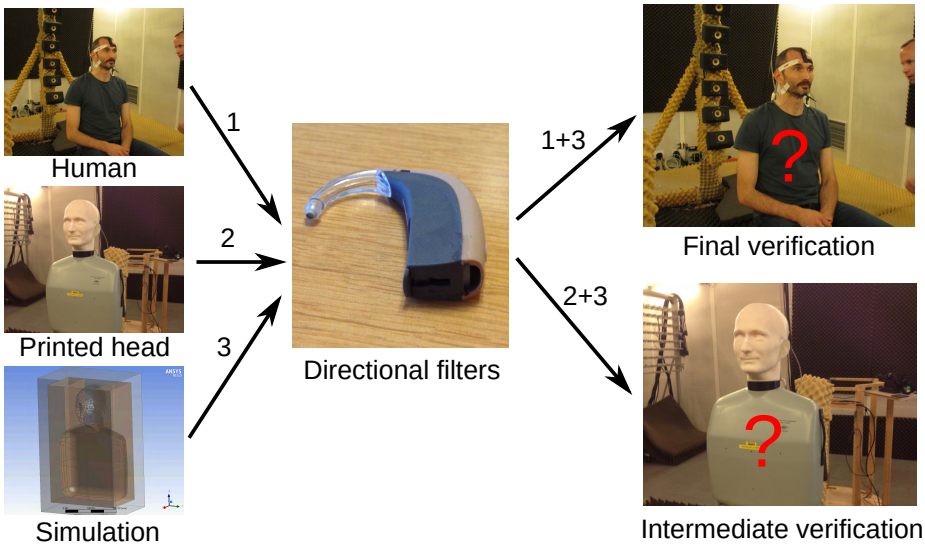


Figure 1.2: Verification of directional filters obtained from either measurement or simulation, using both human and printed head measurements. Numbers indicate the HRTF set used for generating directional filters. 1: HRTFs measured on a human subject, 2: HRTFs measured on a printed 3D head model, and 3: HRTFs simulated based on the same 3D head model.

1.3 Reading guidelines

This thesis is intended to be read in chronological order. Depending on the readers prior knowledge on the human hearing, hearing loss, and hearing aids Chapter 2 and Chapter 3 can be skipped, since they mainly contain a review of these topics. Section 3.2, "Hearing aid types", is however recommended for all readers, because our custom-made hearing-aid shells are introduced in this section. Chapter 5, "Estimation of 3D head models based on 2D images", also stands out because none of the results obtained in this chapter have been employed directly elsewhere in the thesis. With a focus on the main results obtained in this thesis, the chapter can be skipped without loss of reading flow. The remaining chapters contain both a review of existing methodology along with an explanation of how we utilize or develop extensions of the existing methods.

As previously stated, this thesis has been a collaborative work between many different partners and the author has not carried out all of the work single-handedly. It has therefore been chosen to explicitly state, in the beginning of

each chapter, who contributed to the work described.

The thesis furthermore encompasses four papers, which are appended as Paper A-D. The papers will be referred to throughout the thesis where appropriate.

1.4 Summary of papers

Here a short summary is provided for each of the papers included in this thesis. The purpose is to give the reader a brief overview of the goal and main contributions of each paper.

Paper A: A three dimensional children head database for acoustical research and development

Goal: Paper A has two goals. The first goal is to develop a method for easy acquisition of accurate 3D human head models, that extends to children. The second goal is to create a database of high-quality 3D head models of both adults and children for acoustical research.

Main contributions: The first main contribution of Paper A is a scanning, stitching and meshing framework for acquisition of complete human heads. A scan protocol was developed along with the framework to assure consistent data collection. The proposed framework is semi-automatic to assure high-quality and controlled surface representations. The second main contribution is a database of high-quality 3D head models including both adults and children. A 3D head model database is already available with the SYMARE database [10], however their head models are acquired from magnetic resonance imaging (MRI) data, which have an insufficient accuracy for scanning surface models of the external ear. Furthermore, MRI scanning of small children would require sedation. Our head models are acquired using a surface scanner which provides accurate surface representations for both adults and children.

Paper B: A pipeline of geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions

Goal: The goal of Paper B is to investigate whether a 3D-printed listener-specific head model can replace measurements on a human listener and whether numerical simulations can replace acoustic measurements.

Main contributions: The first main contribution of Paper B is the proposed pipeline for 3D printing of a complete human head for acoustical measurements. To the best of our knowledge, such a listener-specific 3D printed head has not been produced previously. The second main contribution is a unique set of HRTFs, consisting of HRTFs measured on a human test subject, HRTFs simulated based on a model of the human subject's head and HRTFs measured for a printed version of the head model. This set of 'human', 'printed' and 'simulated' HRTFs are also novel and it enables an evaluation of whether a 3D-printed head model can replace measurements on a human listener and whether numerical simulations can replace acoustic measurements. A final contribution is HRTF measurements for three additional human test subjects to establish a baseline. All HRTF measurements were performed on a microphone placed in the ear canal. The results showed somewhat similar monaural cues between 3D model measurements and simulations. However, less similar results were observed between human measurements and simulations, which was mainly assigned to differences in torso geometries.

Paper C: Reliability in measuring head related transfer functions of hearing aids

Goal: The goal of Paper C is to evaluate the reliability of HRTFs measured on hearing aids.

Main contributions: The main contribution of Paper C is an evaluation of the reliability of HRTFs measured for BTE and CIC hearing aids, using both a 3D printed listener-specific head model and the corresponding human subject. Even though reliability have been studied previously [MSHJ95, WK89], an evaluation of the reliability of HRTF measurements for hearing aids are novel. Paper C evaluates both within-session and across-session repetitions, where the within-session evaluations had a focus on repositioning of hearing aids as well as repositioning of the printed head. The results showed very reliable measurements for a simple repetition, a repositioning of the hearing aids reduced the reliability, whereas an additional repositioning of the printed head did not sub-

stantially reduce the reliability further. HRTFs measured in different sessions showed a larger variability attributed to less controlled measurement conditions and measurements performed on an actual human listener showed more variability than those with a printed head, probably due to small head or body movements.

Paper D: Optimizing hearing-aid directionality from measurements and simulations

Goal: The goal of Paper D is to evaluate the performance of an individually optimized directional microphone based on FEM simulated HRTFs.

Main contributions: The first contribution of Paper D is an optimization procedure of the directional filters in a BTE hearing aid based on the DI. Furthermore, the paper contributes with a weighting scheme that accounts for non-uniformly sampled source positions. The optimization and weighting procedures are used to optimize the directional microphone based on HRTFs simulated using a listener-specific head model. The application of listener-specific directional filters based on simulated HRTFs is novel in itself. Optimized 'simulated' filters are obtained for a single test subject and verified against optimal filters and non-individualized filters. For comparison, optimal and non-individualized filters are evaluated for three additional test subjects. We showed that FEM simulations on a listener-specific 3D head model can provide directional filters that give a relatively high AI-DI for the individual listener. The gain obtained by changing from the directional filters optimized to give the highest AI-DI for an average population, to simulated listener-specific filters was very limited for our test subject, since the average population filter performed very well. We suggest that individuals who differ from the average population, such as children or elderly, might have a larger benefit from having simulated individual filters than our test subject.

Chapter 2

The human hearing and hearing loss

This chapter contains a review of the human auditory system and the problems that occur with a hearing loss. The functions of an ear are accounted for along with the benefits of binaural hearing. Based on the review of the healthy auditory system, hearing impairments are then explained.

It is important to be familiar with the auditory system and the complexities of hearing loss, to be able to understand the functionality of a hearing aid. However, if the reader is familiar with the human hearing and hearing loss, this chapter can be skipped, without loosing any detail or coherency in the remaining report.

The theoretical part on monaural hearing is primarily based on [SST06, JPR+07, EW05]. The theory presented on binaural hearing is based on [CW99] and the theory on hearing loss is based on [Dil12, EW05].

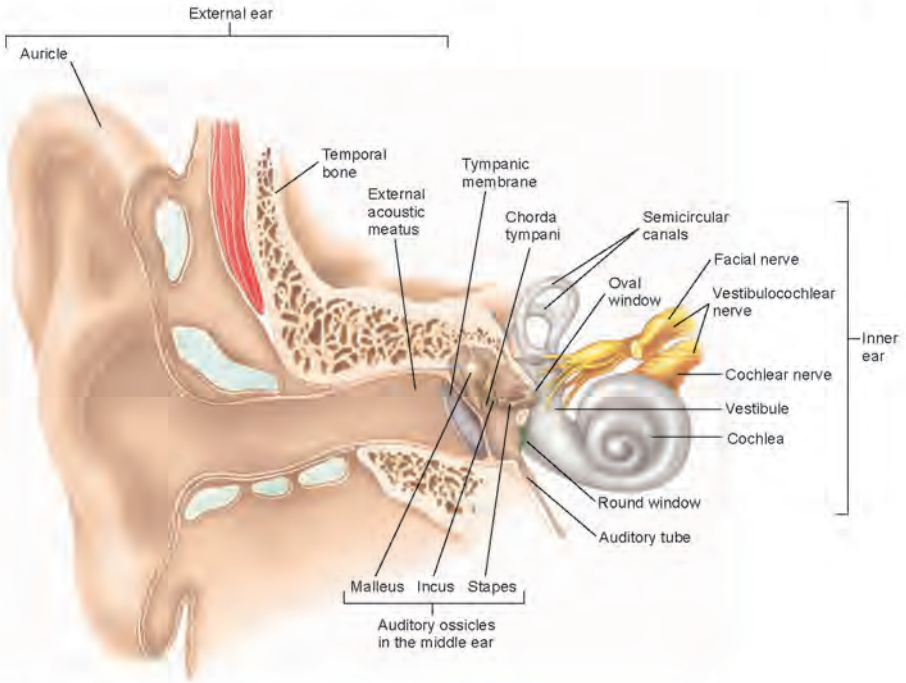


Figure 2.1: The human ear consisting of external, middle and inner ear. From [SST06].

2.1 Monaural hearing

The human ear consists of three parts: the external ear, the middle ear and the inner ear, see Figure 2.1. Each part has an important function for the human hearing and, as Section 2.3 will describe, problems leading to hearing impairment can occur anywhere in the system.

2.1.1 The external ear

The external ear consists of the pinna (the auricle) and the ear canal (external acoustic meatus). Both the pinna and the ear canal are made of elastic cartilage covered with skin. The outer part of the pinna is relatively soft whereas the inner part of the pinna and the ear canal are stiff. The ear canal ends up in the temporal bone just before the tympanic membrane, which defines the transition to the middle ear, see Figure 2.1. The function of the outer ear is primarily to

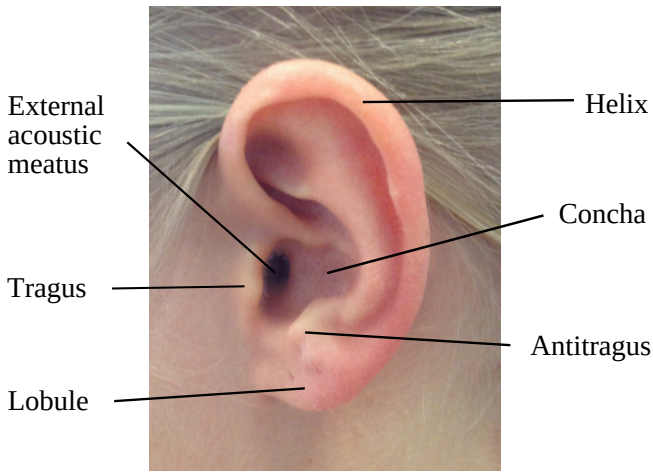


Figure 2.2: The external human ear. The pinna and the outer part of the ear canal are shown.

transfer sounds arriving at the ear canal to the middle ear.

Beside, the pinna also has an essential role for our perception of sounds. First of all the clam-like shape, see Figure 2.2, assures that sounds arriving from the front are captured well, whereas sounds coming from the rear are dampened. Secondly, reflections and diffractions in the pinna changes the sounds that reach the ear canal differently depending on the direction to the sound source. Because our brain is trained to interpret the reflections and diffractions patterns generated by our individual pinna shape, we are able to differentiate between different sound directions.

2.1.2 The middle ear

The tympanic membrane is positioned at the end of the ear canal, on the transition to the middle ear. The middle ear consists of three small bones: hammer (maleus), anvil (incus) and stirrup (stapes). The three structures are seen in Figure 2.3, where they are marked as M, I and S, respectively. The function of these three bones is to transfer vibrations from the tympanic membrane to the inner ear. The hammer receives vibrations from the tympanic membrane, which is then transferred through mechanical motion to the anvil and stirrup. The footplate of the stirrup is connected to the oval window, which is a membrane that connects the air filled middle ear with the fluid filled inner ear. The fluids in the inner ear are incompressible and when the stirrup pushes inwards,

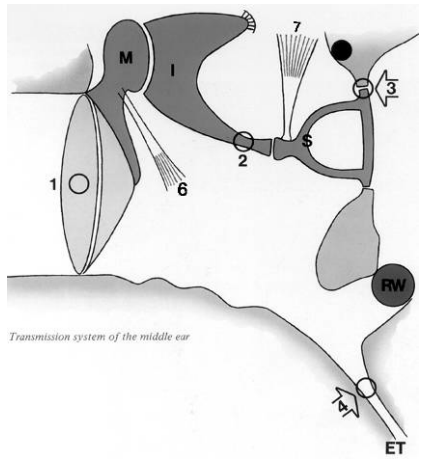


Figure 2.3: The middle ear consisting of the hammer (M), the anvil (I, 2) and the stirrup (S). Furthermore the tympanic membrane (1), the Eustachian tube (ET, 4), the round window (RW) and the two small muscles, tensor tympani (6), and stapedius (7) are marked. (3) marks the point which the footplate rotates around. From [JPR+07].

an equal outwards movement occur at the round window, see Section 2.1.3.

2.1.3 The inner ear

The inner ear consists of the cochlea, the vestibule and the semicircular canals. The vestibule and semicircular canals are primarily related to balance, whereas the cochlea is related to hearing. The main function of the cochlea is to convert vibrations transferred from the stirrup to the inner ear into nerve impulses. The cochlea is a fluid filled snail-shell shaped structure, which is 32 mm long from base (bottom) to apex (top). A cross-section of cochlea is seen in Figure 2.4. The cochlea is divided into three channels: scala vestibuli (1), scala media (2) and scala tympani (3). The oval window, where the sound (mechanical vibration) enters cochlea, is connected to scala vestibuli. The sound (fluid wave) then travels from the base to the apex, where a hole connects scala vestibuli with scala tympani. The sound (fluid wave) travels down scala tympani and exits through the round window.

The basilar membrane (6) divides scala tympani from scala media. The basilar membrane has a thickness that changes with position from apex to base. It is

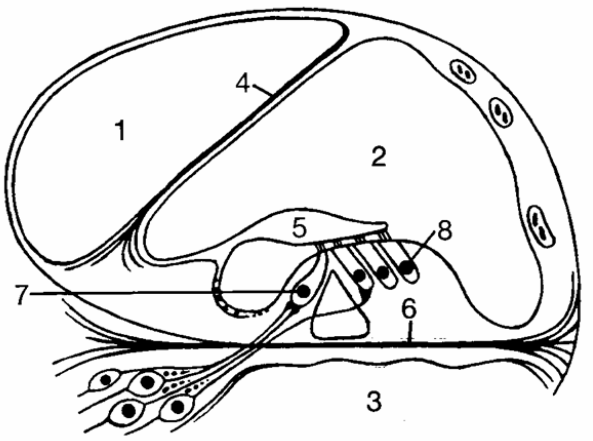


Figure 2.4: Structures of Cochlea. The figure shows scala vestibuli (1), scala media (2), scala tympani (3), membrane (4), organ of Corti (5), basilar membrane (6), inner hair cells (7), and outer hair cells (8). From [JPR+07].

0.5 mm thick at the apex, but only 0.1 mm at the base. The organ of Corti (5) is positioned on top of the basilar membrane. It consists of two sets of hair cells: the inner (7) and the outer hair cells (8). The inner hair cells are the sensory cells. When a sound travels through cochlea, the basilar membrane will vibrate causing the hairs to bend back and forth. This will trigger the inner hair cells to transmit nerve impulses to the hearing center in the brain. The outer hair cells will amplify the vibration of the basilar membrane for weak sounds, so that the inner hair cells react. The basilar membrane works as a frequency analyzer. High frequency pure tones will trigger the basilar membrane near the base, whereas low frequency pure tones will trigger the basilar membrane near the apex. In this way different nerve signals are send to the brain depending on the frequency of the sound.

In summary, the external ear transfers sound to the tympanic membrane. Vibrations in the tympanic membrane are mechanically transferred through the middle ear. The mechanical movements are converted, through the oval window, to pressure changes in the fluid filled cochlea. The pressure changes are here converted to nerve impulses, that are interpreted in the brain. In fact the brain receives nerve impulses from both of our ears and it interprets differences between these two signals to provide the benefits of binaural hearing.

In relation to the remaining of this thesis the pinna functions and the functions of the inner ear are particularly important.

2.2 Binaural hearing

Humans have a binaural hearing, with the two ears placed symmetrically on the left and right side of the head. The binaural hearing contributes with the ability to localize a sound source. Exactly which cues the human brain interprets to estimate a sound source position is not fully understood and it is of great research interest. The Duplex theory by L. Rayleigh [Ray07] can be used to explain directional hearing to some extend.

2.2.1 Duplex theory

Duplex theory is a simple model that explains directional hearing based on interaural time differences and interaural intensity differences. The model is however a simplification of how the brain actually works and Duplex theory is not capable of explaining directional hearing for all 3D positions. It is however an intuitive model.

A sound traveling through air will initially reach the ear that is closest to the sound source. The sound will then reach the other ear, a small amount of time later. The time span between these two occurrences is defined as the interaural time difference (ITD), which is translated by the brain into source positions. The ITDs are only valid for frequencies below 1500 Hz, because wavelengths which are smaller than the head size will lead to aliasing.

Interaural intensity differences (IIDs) are a phenomenon which occur for sounds above 1500 Hz. When the wavelength of the sound is smaller than the size of the head, the head starts to shadow the contralateral ear[1], so that the sound levels received at the two ears differ.

If Rayleigh's Duplex theory is used to model the sound source position in free space, with different azimuth (rotation in the horizontal plane) and elevation (rotation in the median plane) angles, one will notice that the model has some ambiguities where the model does not specify a specific source position. Instead it specifies a cone of positions located on a plane that is parallel to the median plane, this cone is called the *cone of confusion*. Sound sources positioned on the median plane itself is also ambiguous, since the ITD is zero and the IID is zero for all positions.

Humans are, contrary to Duplex theory, able to distinguish between sound sources placed on the cone of confusion and in the median plane. It is be-

[1] The ear opposite to the sound source.

lieved that monaural cues from pinna along with head movements are used to make this distinction [Bla83].

A simple model that fully describes the spatial hearing does not exist. Instead measured head related transfer functions can be used to describe the cues used for human spatial hearing.

2.2.2 Head related transfer functions

Head related transfer functions (HRTFs) are numerical representations of spatial acoustic cues. A HRTF is the transfer function from a point in space to a microphone positioned in the ear canal normalized with a free-field response. HRTFs capture spectral cues from the head, ears (pinna) and torso. They are therefore unique for the individual person [WK97], as well as for the left and right ear. HRTFs are normally obtained for each ear individually and a set of source positions on a sphere surrounding the head. Note that both ITDs and IIDs are contained in the HRTFs. HRTFs have great practical use for instance for generation of 3D sounds or for obtaining directionality. Chapter 8 explains how HRTFs are traditionally obtained through acoustical measurements.

2.3 Hearing loss

The human hearing is, as it has just been described, quite amazing. The functionalities in the different parts of the ear can however quite easily be damaged. Especially the fine hair cells in cochlea are very sensitive. Damage to any part of the ear can lead to a more or less profound hearing loss. Even though the binaural hearing that was just explained is not damaged directly, it will also be affected by the reduced amount of nerve signals to the brain from the two ears.

A hearing loss occurs because of loss of functionality in a part of the audible system. If the problematic area is located before the sound reaches the inner ear, the hearing loss is called *conductive*. If the problem, on the other hand, is located within the inner ear the hearing loss is called *sensorineural*. Examples of a conductive hearing loss could be ear wax that blocks the ear canal, inflammation of the middle ear, a ruptured tympanic membrane or 'mechanical' problems in the middle ear bone chain. A conductive hearing loss gives at most a 50 dB loss of hearing (a severe hearing loss ranges from 70 to 100 dB), and it is often reversible. For a sensorineural hearing loss, the loss of functionality is often caused by damage to the hair cells in cochlea. A small sensorineural hearing

loss involves only damage to the outer hair cells (below 60 dB), whereas a more severe hearing loss (above 60 dB) involves both outer and inner hair cells. A small hearing loss will therefore primarily affect low sound levels, whereas a more severe hearing loss will affect all sound levels. Damages to the hair cells can be caused by noise, medicine, age and other factors.

2.3.1 The audiogram

A hearing loss is often evaluated using an audiogram. The audiogram shows an individuals audible threshold for pure tones of different frequencies. The audible threshold is evaluated for the left and right ear separately. The audiogram has frequency on the horizontal axis and hearing threshold relative to normal hearing on the vertical axis, see Figure 2.5. An audiogram is measured using an audiometer, which presents different pure tones through a headset. The lowest audible sound at a given frequency indicates the hearing threshold. The left ear is indicated with a 'x' and the right ear with an 'o' in the audiogram. A normal hearing individual would have an audiogram that deviates less than 10-15 dB from the zero line. A bone conduction test, where the functions of the inner ear are tested through vibrations in the bone, can provide information about whether the hearing loss is conductive or sensorineural. A bone conduction audiogram is usually marked with a square symbol, see Figure 2.6.

An audiogram for a conductive hearing loss is seen in Figure 2.6a. It is seen that the bone conductive audiogram is normal, whereas the hearing threshold is elevated for the normal hearing test, indicating problems in the ear canal or middle ear. An audiogram for a sensorineural hearing loss is seen in Figure 2.6b. The bone conduction and normal audiogram show similar hearing thresholds, indicating damage in the inner ear.

2.3.2 The five dimensions of hearing

A conductive hearing loss is the most simple kind of hearing loss, because the sound is simply attenuated as it passes through the middle ear. The hearing loss can therefore be improved by a simple amplification of the sound level. The most common kind of hearing loss is however the sensorineural hearing loss, which is much more complicated. In [EW05] a hearing loss is described using the five dimensions of hearing: sensitivity, dynamic range, frequency resolution, time resolution and binaural hearing. Hearing impaired individuals are often affected on all five dimensions of hearing:

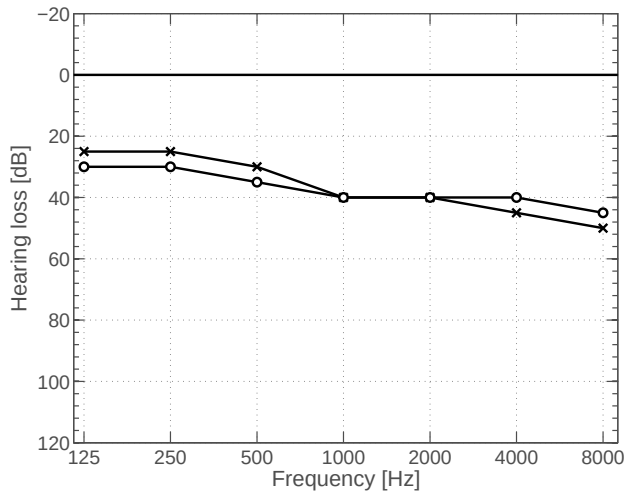


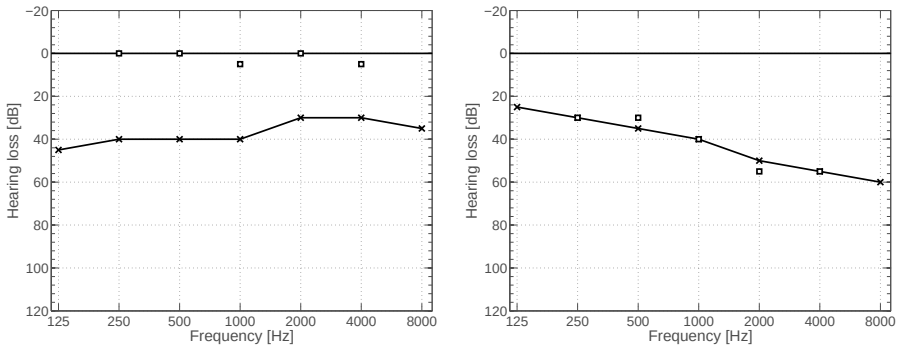
Figure 2.5: Audiogram revealing the amount of hearing loss at different frequencies. Left ear is marked with a 'x' and right ear with an 'o'. Reproduced from [EW05].

Sensitivity: People with a hearing loss have a lower sensitivity towards sounds at different frequencies. The lower sensitivity is detected by the audiogram.

Dynamic range: The range from the softest audible sound to the loudest sound that the individual can listen to without discomfort is called the *dynamic range*. The threshold for the softest audible sound changes a lot when you have a hearing loss, whereas the threshold of loudness discomfort only change a small amount. This reduces the dynamic range. The consequence is that softer sounds must be amplified more than louder sounds, in order to stay within the reduced dynamic range.

Frequency resolution: Individuals with normal hearing are not always able to differentiate between two pure tones if their frequencies are located close to each other. This phenomena occurs because the vibrations in the basilar membrane created by the low frequency tone extend into the area of the basilar membrane where the higher frequency tone is located. Thereby, the low frequency tone tend to cover the higher frequency tone. The ability to differentiate between two simultaneous pure tones is called frequency resolution. Hearing impaired individuals have a reduced frequency resolution, which reduces speech intelligibility, especially in conditions with a lot of background noise.

Temporal resolution: Individuals with normal hearing are not always able to



(a) Conductive **(b)** Sensorineural

Figure 2.6: Bone conductive audiograms for conductive and sensorineural hearing losses, respectively. Reproduced from [EW05].

differentiate between two sounds if they are played shortly after one another. Weak sounds does in particular have a tendency to be covered by more powerful sounds. The ability to differentiate between two subsequent sounds is called the temporal resolution. The temporal resolution is often reduced for hearing impaired individuals and it affects speech intelligibility. Speech consists of many different elements such as consecutive powerful vowel sounds and weak consonant sounds. With a reduced temporal resolution speech will be masked (one sound covering another sound) to a higher degree than for a normal-hearing individual.

Binaural hearing: The binaural hearing is especially important for sound localization and for dealing with situations with a lot of background noise (commonly referred to, as the cocktail-party-problem). The brain of a hearing-impaired individual only receives nerve impulses from some of the sounds that arrive at the ear and this lack of information to the brain, has an adverse effect on the binaural hearing. The reduced amount of information to the brain, makes it difficult for the brain to create the effects of binaural hearing.

A hearing impaired individual often has problems with all of the five dimensions of hearing. A hearing aid is therefore constructed not only to amplify sounds, but also to compensate for all of the aforementioned problems in the best possible way.

Chapter 3

Hearing aids

This chapter describes hearing aid functionality and types of hearing aids. The main part of the chapter is a review and do not explain any new work, however notice that Section 3.2 introduces the hearing aid shells which were manufactured for the measurements performed for this thesis. The theoretical part is based on [Dil12, EW05].

3.1 Hearing aid functionality

Chapter 2 covered the five dimensions of hearing (sensitivity, dynamic range, frequency resolution, temporal resolution and binaural hearing), which is adversely affected for most hearing impaired individuals. A hearing aid is constructed to compensate in the best possible way for a hearing loss, and it therefore consists of parts that attempt to rectify problems with the five dimensions of hearing.

Sensitivity: All hearing impaired individuals have a lower sensitivity towards sounds. A hearing aids primary function is therefore to amplify sounds arriving at the ear canal. The basic components needed for an amplification of sounds are a *microphone* at the entrance, an *amplifier*, a *loudspeaker* at the exit and a *battery*, see Figure 3.1.

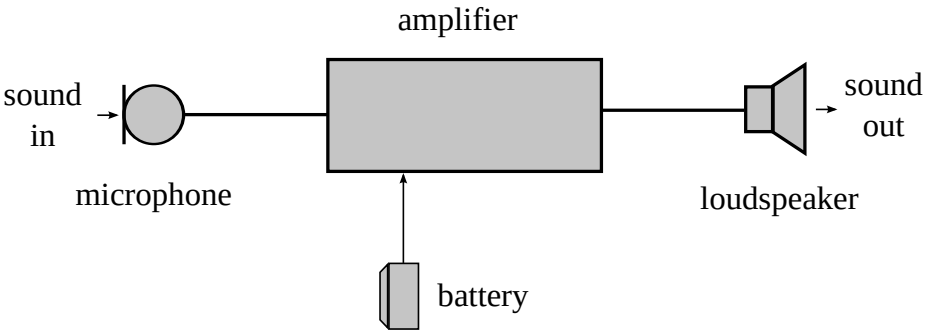


Figure 3.1: The fundamental structure of a hearing aid. The main constituents are microphone, amplifier, loudspeaker, and battery. Reproduced from on [EW05].

The microphone records sounds arriving at the entrance and converts the sound into electrical signal. The amplifier amplifies the electric signal, and the loudspeaker converts the electric signal into sound, which is released into the ear canal of the hearing-aid user. The battery delivers electrical energy to the hearing aid. These basis components in a hearing aid handles the reduced sensitivity, however problems with the four remaining dimensions of hearing must also be addressed.

Dynamic range: A reduced dynamic range is handled by the *compressor*. The reduced dynamic range entails that soft sounds need a larger amplification than louder sounds, to avoid that sounds become uncomfortably loud. The function of the compressor is to change the gain of the output signal depending on the level of the input signal. Thereby, the compressor assures that sounds are amplified enough, however within the dynamic range.

Frequency resolution: The reduced frequency resolution is in particular a problem in situations where speech is mixed with background noise. No methods have so far been developed for hearing aids to separate speech from noise once the sound is inside the hearing aid [Dil12]. The only solution that can ease a reduced frequency resolution is to use helping systems such as a remote transmitter to minimize noise, a directional microphone to reduce sounds from other directions than the front, or a frequency dependent gain, so that lower-frequency parts of speech or noise do not mask the high-frequency parts.

Temporal resolution: A decreased temporal resolution is also most pronounced in situations with noise. The background noise has a tendency to mask the speech signal. Hearing aids cannot completely rectify for a decreased temporal resolution. However, a fast acting compression, where weak sounds

are rapidly increased and intense sounds are rapidly decreased, can make weak sounds more audible. A fast acting compression can only improve the intelligibility a small amount, and the helping systems described for the reduced frequency resolution are also required here.

Binaural hearing: Hearing impaired individuals have a reduced ability to use their binaural hearing, because the input from the two ears to the brain are reduced due to problems with the remaining four dimensions of hearing. A reduced binaural hearing can therefore be improved by improving the other four dimensions of hearing.

Dillon [Dil12] states that only two proven methods exist for increasing intelligibility above that obtainable with a modern well fitted hearing aid. One is to move the microphone closer to the source and the other is to increase the level of direct sound using a directional microphone. It is not always possible or practical to move the microphone closer to the source, which makes the directional microphone highly important for improving intelligibility.

3.1.1 Directional microphones

A directional microphone works by suppressing sounds arriving from some directions, while maintaining good sensitivity from one direction (usually the front). The opposite of a directional microphone is an omni-directional microphone, which has the same sensitivity towards sound from all directions.

The directional microphone in a hearing aid is usually implemented by a first-order subtractive microphone, where the output from two omni-directional microphones are subtracted. A block diagram of a first-order subtractive microphone is seen in Figure 3.2. The block diagram shows that the sound recorded at the rear microphone is time-delayed and then subtracted from the frontal microphone. The time that it takes for the sound to travel between the microphones is called the *external delay*, and it is calculated as the distance between the microphones divided by the speed of sound. The delay that is introduced electrically inside the hearing aid is called the *internal delay*. If the sound arrives from the rear and the internal delay is equal to the external delay, the front and delayed rear signals will cancel each other out and the hearing aid will thus be insensitive to sounds arriving from the rear.

The ratio between the internal and external delay, called the *delay ratio*, r_d, determines the sensitivity to sounds arriving from different directions. The signal at the rear microphone is delayed by multiplying the signal with:

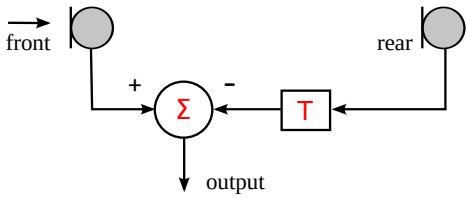


Figure 3.2: Block diagram of a first-order subtractive directional microphone. T is the time delay added to the rear microphone.

$$w_r = \exp\left(-i \cdot 2\pi f \frac{r_d \cdot d}{c}\right), \tag{3.1}$$

where f is the frequency, d is the distance between the front and rear microphone, and c is the speed of sound. The actual delay is $\frac{r_d \cdot d}{c}$.

The sensitivity pattern of a directional microphone is usually indicated using a polar plot. The polar plot reveals the microphones sensitivity to sound arriving from different directions. An example of a polar plot, at a single low frequency, is seen in Figure 3.3. The figure shows a particular polar sensitivity pattern called a hyper-cardioid which occur in free field when the delay ratio is 0.33. We will return to the importance of the hyper-cardioid, here it is simply used to state the terminology used when describing sensitivity patterns. The hyper-cardioid sensitivity pattern has a *main lobe* in the front, with the highest sensitivity towards sound arriving at 0° azimuth. The hyper-cardioid also has a two *directional nulls* placed on the left (approximately 110°) and right side (approximately 250°), respectively. Notice that the directional pattern is symmetrical in free field. The directional nulls induce a lobe in the rear referred to as the *rear lobe*. The size of the main and rear lobes are determined by the position of the directional nulls.

Which directional pattern that is most optimal for a hearing aid changes between listening situations. However, in most real life situations it can be assumed that unwanted noise arrives relatively evenly distributed from all directions. Even if the noise source is placed at a specific direction, room reflections will cause the noise to arrive at the microphone from all directions. On the contrary, the desired sound will often arrive mostly from the front, because the desired sound source is often located closer to the hearing-aid user than the noise sources, leading to more direct sound compared to the amount of reflected sound. A situation where the sound arrives mostly from the front is for instance when the hearing-aid user is listening to speech from an individual standing relatively close. Such a listening situation occurs often for a hearing-aid user and it is

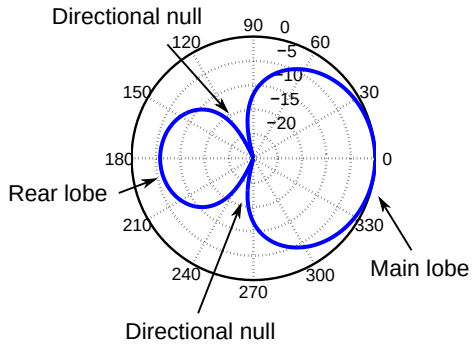


Figure 3.3: Terminology used for describing sensitivity patterns. The shown sensitivity pattern is a hyper-cardioid in free field.

therefore often desired to suppress sounds arriving from all other directions than the front.

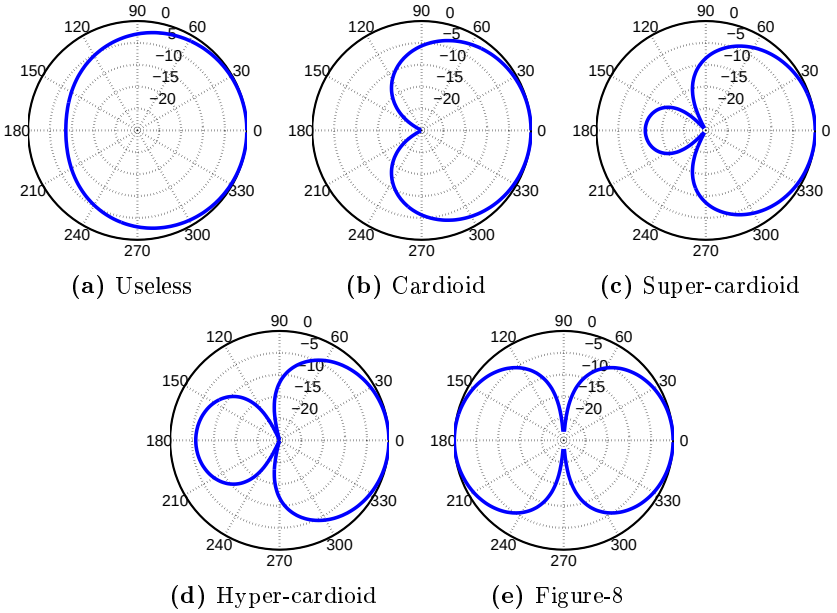
The *directivity index* (DI) describes the sensitivity to sounds arriving from the front relative to an average of sounds arriving from all directions [Dil12]. The DI can be computed for either 2D, using only sound sources in the horizontal plane, or for 3D. Other directionality measures exist, such as the *front-to-back ratio* or *unidirectional index* (UI). The front-to-back ratio is simply the ratio between the front and rear directions and it ignores sounds coming from other directions. The front-to-back ratio is therefore not an appropriate measure for the effectiveness of a directional microphone. The UI describes the sensitivity to sounds arriving from the front region ($-90°$ to $90°$ azimuth) relative to the sounds arriving from the rear region ($90°$ to $270°$ azimuth). The UI is a quite useful measure, however the DI is more commonly used and is therefore the preferred measure throughout this thesis. We will return to directional patterns and DI in Chapter 10.

Sensitivity patterns for delay ratios ranging from 2 to 0 are shown in Table 3.1 with corresponding 2D and 3D DI calculated for microphones placed in free field. Figure 3.4 shows corresponding sensitivity patterns. A delay ratio above 1 does not make sense because the rear microphone is then delayed more than the time that it takes the sound to travel from the front to the rear microphone. Changing the delay ratio from 1 to 0 shifts the directional nulls from a position towards $180°$ azimuth to a position towards $90°$ azimuth. The sensitivity pattern for a delay ratio of 1 has no sensitivity towards sounds arriving from the rear, whereas a delay ratio of 0 results in a sensitivity pattern with equal sensitivity towards front and rear, however no sensitivity towards sounds arriving from the sides. Note that these easily interpreted sensitivity patterns are only obtained in free field. The presence of a head will change the patterns due to reflections

and diffractions of the sound.

	Useless	Cardioid	Super-cardioid	Hyper-cardioid	Figure-8
2D DI	3.0	4.3	4.8	4.6	3.0
3D DI	3.2	4.8	5.7	6.0	4.8
Delay ratio	2.00	1.00	0.577	0.333	0.000

Table 3.1: Free field 2D DI and 3D DI for varying delay ratios. The corresponding sensitivity patterns are seen in Figure 3.4.



(a) Useless **(b)** Cardioid **(c)** Super-cardioid

(d) Hyper-cardioid **(e)** Figure-8

Figure 3.4: Free-field sensitivity patterns obtained for varying delay ratios. Front is positioned at 0° azimuth.

3.2 Hearing aid types

Many different hearing aid models exist and we use some of the standard types during this thesis. For the work performed for the present thesis, the microphone position and the number of microphones in the hearing aid are important. The microphone position is important because the sound field might differ for different hearing aid positions. The number of microphones is important because

a directional microphone can only be implemented if the hearing aid contains two omni-directional microphones instead of only a single microphone.

Hearing aids can be divided into several types according to the hearing aid's position on the ear: Receiver-In-The-Ear (RITE), Behind-The-Ear (BTE), In-The-Ear (ITE), In-The-Canal (ITC), and Completely-In-Canal (CIC).

The BTE hearing aid is positioned behind the ear and a plastic tube leads the sound into the ear canal. The plastic tube is fixed in the ear canal with a dome. This kind of hearing aid is mass-produced, while the size of the dome is individually fitted. The position of the BTE hearing aid is unfortunate with respect to maintaining the great properties that pinna has. However, the position behind the ear is beneficial due to more space, less feedback problems and less need for custom made production. BTE hearing aids are equipped with two microphones and signal processing that can compensate for some of the lost pinna properties.

The ITE hearing aid is positioned partly in the ear canal and partly in the part of the outer ear called concha. The ITE hearing aid is to a large degree custom-made, since the shell of the hearing aid is produced based on an ear impression. The hardware inside the hearing aid is mass-produced, however the positioning of the hardware is adjusted to the individual shell-shape.

The ITC hearing aid is comparable to an ITE hearing aid. However, it is smaller and positioned mainly in the ear canal with only a small part of the hearing aid located in the outer ear. The ITC is produced in a similar fashion to the ITE.

The CIC hearing aid is a very small hearing aid that is placed deep into the ear canal. The CIC shell is also based on an ear impression and hardware space is in particular a challenge for the small CIC hearing aid.

We use hearing-aid shells with microphones to achieve a microphone position compatible with the microphone position that hearing-aid users have. BTE, ITE, and CIC shells were manufactured at Oticon A/S, see Figure 3.5. The shells were equipped with microphones and connectors. The ITE and CIC shells were custom-made based on ear impressions from three test subjects. The BTE and ITE shells contain two matched microphones, whereas the CIC shell only contain a single microphone. Notice, that the produced ITE shells are small compared to a normal ITE. The position of the ITE hearing aids are therefore deep in the ear and resembles the position of an ITC hearing aid.

The ITE and BTE hearing-aid shells are the most interesting types of hearing aids for the work performed for this thesis, because both hearing-aid types contain two omni-directional microphones, which can be combined into a directional microphone. The directional microphone in both types of hearing aids

will be optimized in Chapter 10.

(a) Behind-The-Ear. **(b)** In-The-Ear. **(c)** Completely-In-Canal.

Figure 3.5: Hearing aid shells manufactured for our acoustical measurements.

Chapter 4

A 3D head and ear database

The focus of this chapter is to explain the scanning, stitching and meshing method that we have used to obtain listener-specific head models, and to introduce the 'DTU Compute 3D head model database', which has been established during the project. The scanning, stitching and meshing method described here is also presented in Paper A.

The work presented in this chapter consists of both method development, data collection and many hours of data postprocessing. The scanning protocol, seen in Appendix A.1, was developed by Rasmus R. Paulsen and the author. The scanning, stitching and meshing method presented is an adaption from the surface alignment and reconstruction method developed by Rasmus R. Paulsen et al. [PL10]. The DTU Compute 3D head database in Section 4.5, consists of data from many surface-scan sessions, which were primarily conducted by the author. The postprocessing of the collected data was also, to a great extend, performed by the author. However, Marie E. K. Nielsen, on behalf of GN Resound, carried out data collection and postprocessing for a smaller number of test subjects.

4.1 Listener-specific 3D head models

The focus of this thesis is listener-specific optimization of the directional microphone in a hearing aid. Directionality is optimized on the basis of listener-specific HRTFs, which can be obtained from acoustical measurements. However, such measurements are very time-consuming, cumbersome and require expensive equipment to perform, see Chapter 8. It is therefore of great interest to replace HRTF measurements with HRTF modeling. Our approach is to obtain listener-specific HRTFs through finite element method (FEM) simulations, which is the topic of Chapter 7. In order to simulate listener-specific HRTFs with FEM, a listener-specific 3D head model is required. This section explains our method for obtaining such models.

With a broader perspective than the goal of this thesis, individual 3D head models have many applications. Virtual product design is an emerging discipline, and our listener-specific head models could form the basis for a statistical head and ear model. An example of a statistical model within virtual product design could be the size-China project where population statistics on human heads were created for product design [LBJ12]. Human statistics were also used in [BL12] for creating a parametric model of the entire human body based on 250 full body scans. However, the ears were not captured with sufficient accuracy in [LBJ12] for our purpose.

Listener-specific head models, and certainly also a statistical head model, could be used for virtual product design within the hearing aid industry. It is obvious that virtual product design is beneficial for designing the outer casing. However much more important is the acoustical properties of the hearing aid. The position of the microphones in a hearing aid is essential for capturing spatial acoustic cues, which affects the hearing aid sound. Therefore, different types of microphones at various positions have been proposed [MVCL12]. Our head models facilitate that different microphone positions can be tested using acoustical simulations, it is however outside the scope of this thesis.

4.2 3D scanning of a head

A large range of different scanners exist that are able to capture the 3D shape of a head: laser scanners, structured light scanners, stereo-photogrammetry scanners and medical scanners, such as magnetic resonance imaging (MRI) or X-ray computed tomography (CT). Not all of the scanners are however appropriate for scanning a head with the purpose of generating listener-specific head models.

The MRI scanner has a poor soft tissue resolution [ASM+98], which is not sufficient for accurately scanning a surface model of the external ear. Furthermore, MRI scanning requires that the test subject lays still for a longer period of time. Scanning of a younger child would therefore require sedation. The CT scanner involves radiation and is therefore not an appropriate scanning procedure. Additionally, both MRI and CT scanners are highly expensive and not easily accessible.

Surface scanning an entire head using a laser scanner would require a system with a laser scanner and a rotating chair as the system described by Moss et al. [MLGM89]. Moss et al. have shown the value of such a system, however the image capture is slow (15-30 s) and might lead to motion artefacts when scanning live test subjects [FDGR11].

Instead, we use 3D stereo-photogrammetry to capture 3D head scans. 3D stereo-photo-grammetry, also called stereo vision, is a method that, given two or more images and the relative geometry between the cameras, derives depth information from the scene. Stereo vision usually consists of camera calibration, establishing point correspondences between the images and 3D reconstruction of points in the scene [SHB08]. 3D stereo-photogrammetry overcomes the motion artifacts that the laser scanner induces, by having a near-instantaneous image capture. Furthermore, it has the required resolution for capturing the structures of the outer ear. 3D stereo-photogrammetry has the limitation that it cannot scan occluded areas. The outer ear contains many occluded areas due to its complex structure. Scans from several different angles must, therefore, be acquired and merged together in a postprocessing step.

4.2.1 Our scan setup

We used a Canfield scientific Vectra M3 commercial surface scanner, see Figure 4.1. The surface scanner has an accuracy of approximately 0.2 mm and each scan consists of around 65.000 vertices and 130.000 triangles. Surface textures are computed by the scanner firmware along with the surface. Notice in Figure 4.1 that the test subject is wearing a hair covering device, because the surface scanner cannot scan hair and beard.

A single scan only captures the non-occluded areas of the head. That means that a single scan only covers one side of the head, and that multiple scans are needed in order to achieve data covering the entire head. We developed a scan protocol to assure consistent data acquisition from all angles. The protocol is found in Appendix A.1. In the protocol eight scans are required from different angles surrounding the head. The pinna region is very difficult to scan properly

Figure 4.1: Scan setup with a Canfield scientific Vectra M3 surface scanner.

using a surface scanner, because the pinna's complicated structure gives rise to many occluded areas. The scan protocol was therefore developed with a focus on pinna, and an additional eight pinna scans were captured, where the head was tilted in different directions. The protocol thereby requires 16 surface scans in total.

Having the 16 scans, the goal is to combine the scans into a complete surface model of the head. Combining multiple scans requires surface alignment and surface reconstruction.

4.3 Surface alignment and reconstruction

Different methods exist for surface alignment and reconstruction. Many of the methods are based on the iterative closest point (ICP) algorithm for the alignment [TL94, LPC+00, PL10], however they employ different approaches for reconstruction. Turk and Levoy [TL94] use ICP followed by a zippering of triangle meshes, and therefore handle alignment and reconstruction separately. Curless and Levoy [CL96] present a method using a cumulative weighted signed distance function to create an implicit isosurface that can be extracted as the final mesh. An implicit isosurface approach is also used in Paulsen et al. [PBL10]. Here a markov random field (MRF) based regularization is applied due to the useful hole filling and noise handling properties. The MRF method reconstructs missing parts in an anatomically plausible way, which is very useful for recon-

structing models of anatomical parts such as the head and ears. The method was extended in [PL10] to include a groupwise alignment of the surface scans. We use an adapted version of this method for the reconstruction of our 3D head models. The method will be explained in Section 4.3.1.

4.3.1 Stitching and meshing our head models

After the scan procedure, described in the scan protocol, our goal is to combine the surface scans, captured from different angles surrounding a test subject, into a complete surface model of the head. A conceptual description of the method is given here, whereas a complete description of the method is found in Paper A.

The major difficulty, when combining scans, is the large shape variation among a set of scans. Even though a single scan is near instantaneous, it is impossible to avoid changes in facial expression and head position between scans. The shape variation is in particular a problem when scanning small children, since they cannot be expected to stay in the same position during an entire scan session, see Figure 4.2.

Figure 4.2: Three surface scans of a child, obtained during the same scan session. The figure shows the large shape variation that must be expected when scanning a young child. Take note of the mouth area and head position, for instance.

The shape variation among a set of scans is handled by assembling the final model of a number of sub-parts. The number of scans used for each sub-part is determined by the consistency of the sub-part. A consistent part is here defined as a sub-part with little variation between the set of scans. A good example of a consistent part is the pinna that normally does not deform even with changing

head pose and facial expression. The facial and neck regions normally vary significantly between scans.

Six sub-parts are manually annotated with scaler values: face (1), left ear (2), right ear (3), top of the skull (4), bottom of the skull (5), and neck (6). The set of sub-part scans, that the final model should be assembled from, are manually selected based on the consistency of the sub-parts. Several scans are used for consistent parts such as the pinna, whereas only a single scan is used for the facial area.

The set of partial scans is rigidly aligned using 93 manually annotated anatomical landmark points on the face and ear regions. All scans are aligned to one reference scan. The resulting point cloud has areas that are well covered with points, however it also has areas that lack data points, see Figure 4.3. The point cloud also contains gaussian noise and outliers. Fortunately, the five step iterative approach described in [PL10] handles holes and noise well, and we utilize the method directly.

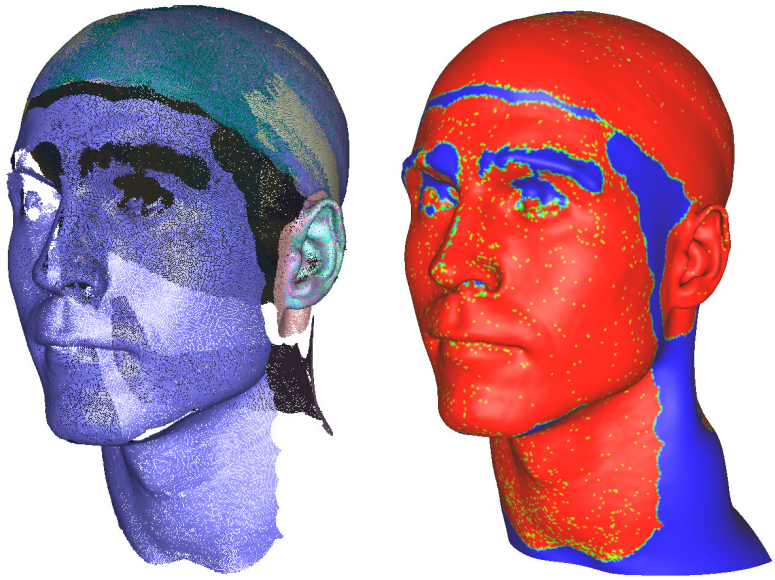


Figure 4.3: Left: Roughly aligned point clouds. Right: The resulting surface visualizing the confidence map where red is high confidence and blue no confidence.

The five step iterative approach consists of:

Step 1: Point Set Merging Input point sets are merged into one point set and point normals from the scanner are extracted.

Step 2: Computing the Signed Distance Field A signed distance field is computed. The distance field is represented as a voxel volume with isotropic voxel size. The distance in a voxel is computed as the distance from the center of the voxel to the average of the five closest points.

Step 3: Markov Random Field Regularization The distance values and signs in the signed distance field can be considered accurate when the distance and sign are sampled close to a dense part of the input point cloud. However, in regions where there is missing data or holes in the point cloud, the confidence in the distance and sign should be very small. To create a consistent distance field a regularization of the voxel values in the distance field is performed. The regularization is formulated in a MRF framework due to the obvious neighborhood relation between voxel values. In the MRF framework two energy terms are formulated. The first energy function, called the prior model, defines how the voxel values would behave if no point data where available. Experiments have shown that using the difference between neighboring Laplacians:

$$U_{\mathcal{L}}(d_i) = \sum_{i \sim j} (\mathcal{L}(d_i) - \mathcal{L}(d_j))^2 \tag{4.1}$$

enforces a desirable higher order behavior of the iso-surfaces of the distance field. Details of this behavior is explained in detail in [PBL10].

To be true to the original point cloud an observation model is also formulated as an energy term. This is a simple quadratic term that penalize that the distance value in a voxel close to the point cloud is changed to much in the regularization:

$$U_{\text{obs}}(d_i^o) = (d_i - d_i^o)^2, \tag{4.2}$$

where d_i^o is the original distance at voxel i and d_i is the current estimate.

Finally, a local weight α_i is used to balance the prior and observation model. α_i is locally computed as the confidence in the local point part of the point cloud. It is close to one when a voxel is close to a well defined part of the point cloud and close to zero when the voxel is far away from well defined points.

The overall goal is therefore to recompute a signed distance field which in our case is a voxel volume that maximizes the posterior probability:

$$\hat{\mathbf{d}} = \arg\max_{\mathbf{d}} \; p(\mathbf{d}|\mathbf{d}^o), \tag{4.3}$$

where the Gibbs posteriori probability per voxel becomes:

$$p(d_i|d_j, i \sim j) = \exp\left(-\alpha_i\beta\ U_{\text{obs}}(d_i) - (1 - \alpha_i\beta)\ U_{\mathcal{L}}(d_i)\right). \tag{4.4}$$

The global weight β can be used to control the overall influence of the prior contra the observation model. In practice, tuning β effects the overall smoothness of the result.

The solution to Equation 4.3 is found by a multi-threaded multi-scale iterative conditional modes solver. Details can be found in in [PBL10]. The result is a regularized distance field with an anatomically plausible behavior of the zero-level iso-surface.

Step 4: Realignment The individual point sets are realigned to the zero-level iso-surface using an implicit version of ICP. Since this iso-surface is both smoothed and regularized, the influence of the individual point sets have been blended together.

Step 1-4 is iterated, with a decreasing scale and thereby increasing resolution.

Step 5: Surface Extraction A polygonized version of the isosurface is extracted using marching cubes [LC87] and the mesh is the optimized for a more regular mesh.

4.4 Evaluation of our 3D head models

The accuracy of the 3D head and ear models obtained here is of great interest since it influences the acoustical simulations and thereby the final optimization of the hearing aids. However, it is not straightforward to calculate the accuracy since we are missing a golden standard in the form of a correct geometry of our test subject's head. A golden standard could for instance be a 3D head scan obtained using a different scan modality, however the accuracy of such data is also unknown.

For the purpose of providing a golden standard, G.R.A.S. Sound and Vibration has kindly provided us with the geometry of KEMAR. This geometry works as a golden standard for the 3D head model of KEMAR. The geometry forms a great opportunity for testing and evaluating the scanning, stitching and meshing process. It is even possible to get an idea of which areas that are problematic to scan and reconstruct.

We utilize the scanning, stitching and meshing approach just described to generate a 3D model of KEMAR. A comparison of the reconstructed 3D model and the geometry of KEMAR can be seen in Figure 4.4. The accuracy is quite good in the facial and upper head regions. Large inaccuracies are observed in the back of the head, and they arise because KEMAR was equipped with a wig cap to reduce scan artifacts in the area, adding a layer which are not present in the virtual geometry. Furthermore, the neck area contains large inaccuracies. In the areas most relevant for us, with respect to acoustical modeling, the largest errors are located in the pinna regions. The complicated structure of the pinna makes it difficult to scan properly. Some areas will remain occluded even though we seek to avoid it by capturing additional scans from different angles. These areas contain no data and the reconstruction process has to 'guess' the shape of the surface, leading to a more inaccurate result. One particularly problematic area is the back side of the left pinna, marked in Figure 4.4, which contains errors above 2 mm. Besides this area, errors in the pinna regions seem to be below 2 mm in most areas, which we consider a successful result.

Even though a comparison between the KEMAR geometry and model is very informative, there is some differences between scanning KEMAR and a normal human. One advantage is that KEMAR has no shape variation between the different scans. Obviously, both facial expression and head remain unchanged. A disadvantage when scanning KEMAR is that the surfaces are very uniform and shiny, which leads to more scan artifacts because of reflections from the flash. A surface scan of KEMAR is seen in Figure 4.5. Notice the many scan artifacts especially in the neck area. Fortunately, we are less concerned with the neck area.

Using the KEMAR geometry as a golden standard, has provided valuable information about the accuracy of our head models. Because of the mentioned differences between scanning the KEMAR and a human test subject, exactly the same accuracy cannot be assumed for our human models. We do however believe that the largest errors in a human model are located in the pinna areas because of the problem with occlusion, which was observed for the KEMAR. Furthermore, we believe that the maximum errors will be in the 1-3 mm range as observed for the KEMAR. The accuracy of a printed version of a human head model is accessed in Chapter 6.

4.5 The DTU Compute 3D head database

We have created a database of 3D head models using the scanning, stitching and meshing method described above and in Paper A.

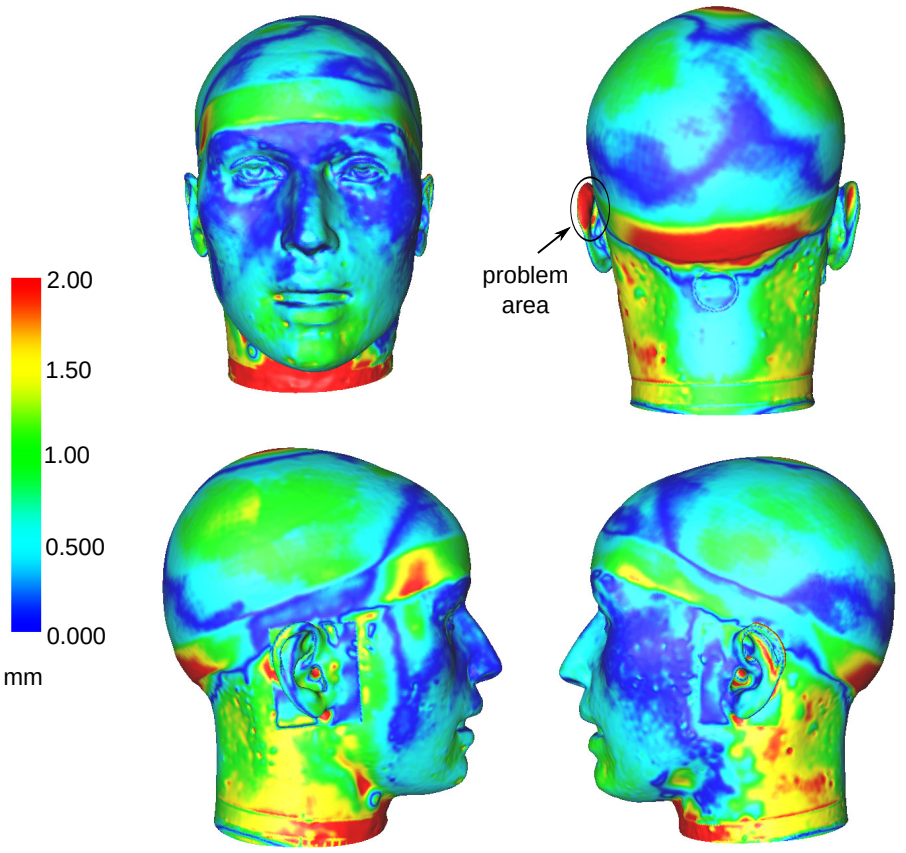


Figure 4.4: Error map between the KEMAR geometry and the reconstructed 3D model of KEMAR.

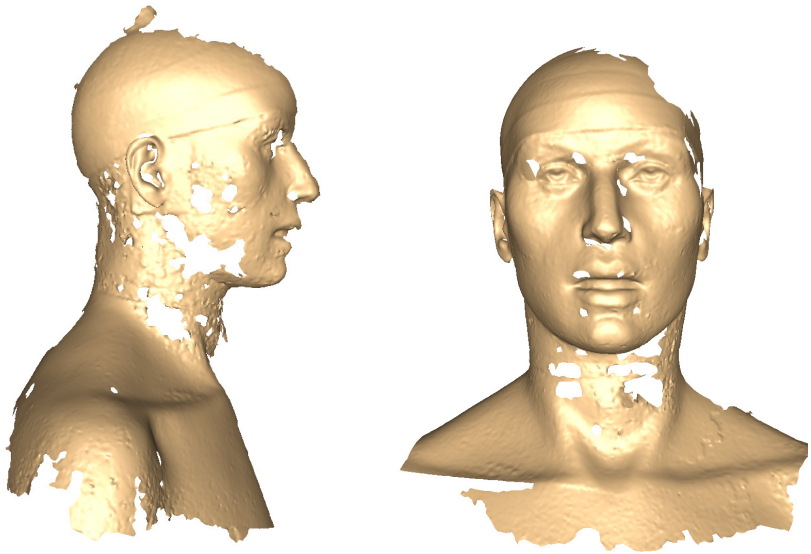


Figure 4.5: 3D surface scans of KEMAR. The amount of scan artifacts is high especially in the neck area.

The data was collected using a Canfield scientific Vectra M3 surface scanner, sponsored by grants to Thomas Werge from The Danish National Advanced Technology foundation and Lundbeck Foundation. The scanner was, during the scan period, placed at three different locations: Distriktspsykiatrisk Center Østerbro, Enheden for Psykiatrisk Forskning at Aalborg Psykiatriske Sygehus and the Technical University of Denmark.

The database currently contains reconstructed 3D head models of 30 adults and 4 children, see Figure 4.6. In addition, surface scans are available for reconstruction of approximately 35 more head models, leading to a total of approximately 70 scanned test subjects.

The aim of this database is to generate a publicly available database of high-resolution 3D scans of the head and ears. The 3D head models are intended for acoustical modeling or virtual design of for instance hearing aids or headsets. However, many other applications probably exist. Throughout this thesis we use the 3D model of the subject in the upper left corner of Figure 4.6, which we from now on will refer to as subject NH167. The head model of subject NH167 is used both for 3D printing (Chapter 6) and for acoustical modeling of head related transfer functions (Chapter 7).

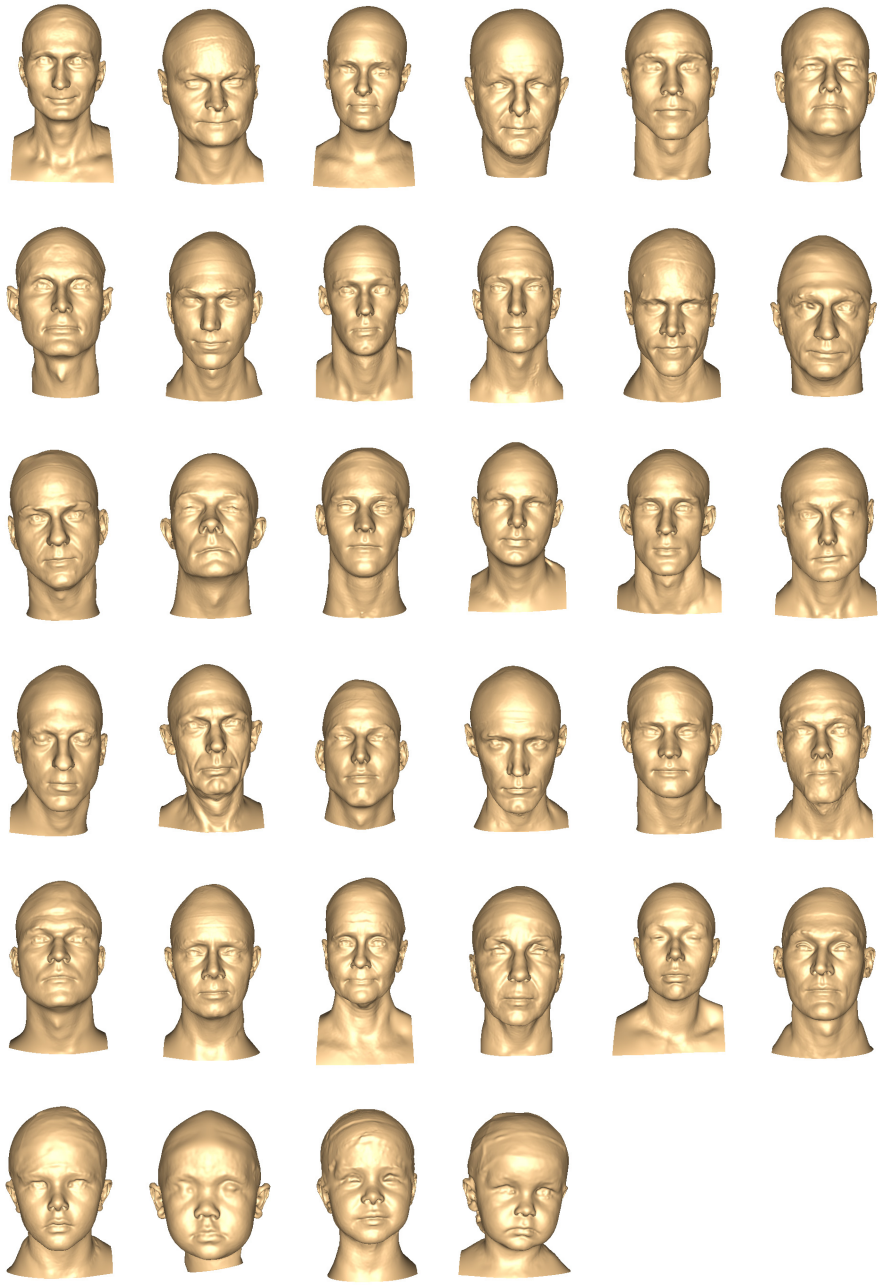


Figure 4.6: DTU Compute 3D head database currently consisting of 30 adults and 4 children.

Chapter 5

Estimation of 3D head models based on 2D images

This chapter explains already existing methods for building a statistical 3D shape and texture model, also called a 3D morphable model (3DMM) for faces and for estimating 3D faces based on 2D face images. We utilize the existing methods to build our own 3DMM based on a large database of face scans. Furthermore, we extend the method in order to build a statistical 3D ear model. We foresee that the presented statistical shape models along with 2D images can be used to generate listener-specific 3D head models for individual HRTF modeling. However, our work on the statistical shape models and on fitting the shape models to 2D images are currently not at stage to be readily integrated with the HRTF modeling. If the latter is the focus of the reader, this chapter can therefore optionally be skipped, without any significant loss of information in the later chapters.

The work described in the present chapter consists of two parts. The first part on generating a 3DMM for faces and fitting the 3DMM to a 2D image were made in collaboration with Jens Fagertun and Rasmus R. Paulsen. The correspondences for the facial 3DMM were generated by Rasmus R. Paulsen and the model was built by Jens Fagertun. Fitting the facial 3DMM to a 2D image was done in collaboration with Jens Fagertun. The second part on developing a statistical model of the outer ear was made in collaboration with Professor

Serge J. Belongie from University of California, San Diego. Implementations were made by the author.

5.1 Estimating 3D models based on 2D images

An ambitious long-term vision is to develop a pipeline where an audiologist uses a mobile phone or a digital camera to acquire images of the customer's head and ears. From these 2D images a corresponding 3D model can be generated. Based on this model, individually optimized directional filters will be computed based on FEM simulated HRTFs. The individual directional filters will be stored in the hearing aid, which is then returned to the audiologist. An important aspect of this pipeline is that the process of capturing 2D images is not too time consuming nor too complicated for the audiologist.

Estimation of 3D structures based on 2D images can be performed using a highly calibrated camera setup [HZ03]. The method is called stereo-photogrammetry and the principle is that depth information from the scene is derived based on two or more images and the relative geometry between the cameras [SHB08]. Even though stereo-photogrammetry is a state-of-the-art method for generating 3D models and works well for generating a 3D head model, current methods and systems require too much calibration of the cameras to be a feasible method in the clinic of an audiologist.

The alternative method, where a strong statistical prior is used to predict structures with known statistical shape priors from one or a few 2D images, is therefore interesting. Since we want to estimate full 3D head models it is in particular interesting, that it has been proven possible to estimate the 3D anatomy of human faces from frontal photos [BV03]. Solving the problem of estimating 3D head models based on a statistical prior and 2D images consists of two major tasks: The first is to build the statistical head model by creating a dense point correspondence between a high number of head models. The second is to fit the statistical shape model to 2D images.

We perform a conceptual trial of the method proposed by Blanz and Vetter [BV03] using 3D scans of the facial region, see Section 5.2. The reason for using facial scans instead of complete head models is that the method was originally developed for faces and that we have a large database of 3D facial scans available. Furthermore, a dense point correspondence is easier obtained for facial geometries than for the complicated head geometries.

Creating a statistical model of an entire head is a different task than creating

a statistical model of faces. Especially the complex structures of pinna complicates the task of creating point correspondences. Section 5.5 therefore pursues the challenge of building a statistical ear model. The overall idea is to create separate statistical ear and head models and to combine them subsequently. The last task remains unsolved so far.

5.2 The 3D morphable model

The 3D morphable model (3DMM), introduced by Blanz and Vetter [BV03], is basically a statistical model of shape and texture. The Basel Face Model [PKA+09], which has been made publicly available is an example of a 3DMM of faces. Because a 3DMM explains the variation of the data set from which is was built, a high number of faces are required to achieve a model that represents the human adult population well.

A 3DMM is built in two steps. First, a dense point correspondence is created among the data set, and secondly a principal component analysis (PCA) [Jol02] is applied to find the largest modes of variation. Blanz and Vetter [BV03] use an optic flow approach to create the point correspondences, but other methods exist [PLN+02].

The registered 3D face scans are represented by a shape and a texture vector [PKA+09]:

$$\begin{aligned} s &= (x_1, y_1, z_1, ..., x_m, y_m, z_m)^T \\ t &= (r_1, g_1, b_1, ..., r_m, g_m, b_m)^T, \end{aligned} \tag{5.1}$$

where (x_m, y_m, z_m) and (r_m, g_m, b_m) are the coordinates and rgb-color values of the mth vertex.

A PCA is performed in order to obtain a number of principal components explaining the largest modes of variance with respect to shape and texture. The result is a parametric face model:

$$\mathcal{M}_s = (\mu_s, \sigma_s, U_s) \text{ and } \mathcal{M}_t = (\mu_t, \sigma_t, U_t), \tag{5.2}$$

where $\mu_{s,t} \in \mathbb{R}^{3m}$ are the means, $\sigma_{s,t} \in \mathbb{R}^{n-1}$ are the standard deviations and $U_{s,t} = [u_1, ..., u_n] \in \mathbb{R}^{3m \times n-1}$ are orthonormal bases of principle components for shape and texture [PKA+09].

The 3DMM is capable of generating new faces[PKA+09]:

$$s(\alpha) = \mu_s + U_s \text{diag}(\sigma_s)\alpha \qquad (5.3)$$
$$t(\beta) = \mu_t + U_t \text{diag}(\sigma_t)\beta,$$

where α and β are weighting parameters that can be used for generation of new faces and model fitting.

We build our 3DMM from 605 3D face scans obtained from the Danish Blood Donor Study [PEK+12]. The face scans were manually annotated with 73 landmarks. We use the method developed by Rasmus R. Paulsen [PLN+02] to create a dense point correspondence. A template face scan is warped onto all of the remaining face scans, thereby creating correspondence between the entire set of scans. The 73 annotated landmark points are used as a sparse correspondence for guiding a Thin Plate Spline (TPS) warp. Once the template has been warped onto another scan a dense correspondence is established by projecting all vertices from the template onto the scan. This is followed by a regularization of the corresponding field. Once the dense point correspondence is created for all samples a PCA is applied to obtain the principal components with the highest variation for shape and texture.

The behavior of the 3DMM can be observed in Figure 5.1 and 5.2. The figures show a mean face in the center position, and the left and right images show minus and plus three standard deviations (SDs) of the first three principal components (PCs) with respect to shape and texture.

The first PC for shape explains the variation in the height of the head, the second PC explains the variation in the roundness of the head, whereas the source of variation in third PC is more unclear, see Figure 5.1. With respect to texture, the first PC explains light versus dark variations, the second PC explains variations with beard or no beard, and the third PC explains if the general color is grayish or reddish, see Figure 5.2. All three texture variations appear to contain a female versus male variation. The first principal component also appear to capture reflections from the flash.

5.3 Fitting a 3DMM to 2D images

A 3DMM is fitted to a 2D image by optimizing the shape and texture coefficients α and β. If a 2D image is given by:

$$I_{input}(x, y) = (I_r(x, y), I_g(x, y), I_b(x, y))^T, \qquad (5.4)$$

(a) First PC.

(b) Second PC.

(c) Third PC.

Figure 5.1: The three PCs with the largest shape variation. Center image: mean face, Left image: mean face minus 3 SDs, Right image: mean face plus 3 SDs.

(a) First PC.

(b) Second PC.

(c) Third PC.

Figure 5.2: The three PCs with the largest texture variation. Center image: mean face, Left image: mean face minus 3 SDs, Right image: mean face plus 3 SDs.

where $I_r(x,y)$, $I_g(x,y)$ and $I_b(x,y)$ are the r, g and b color values at pixel (x,y). Then α and β are optimized by minimizing the sum of squared difference, E_I, between the 2D image and the synthetic reconstruction of the 3DMM for all color channels and all pixels [BV03]. That is:

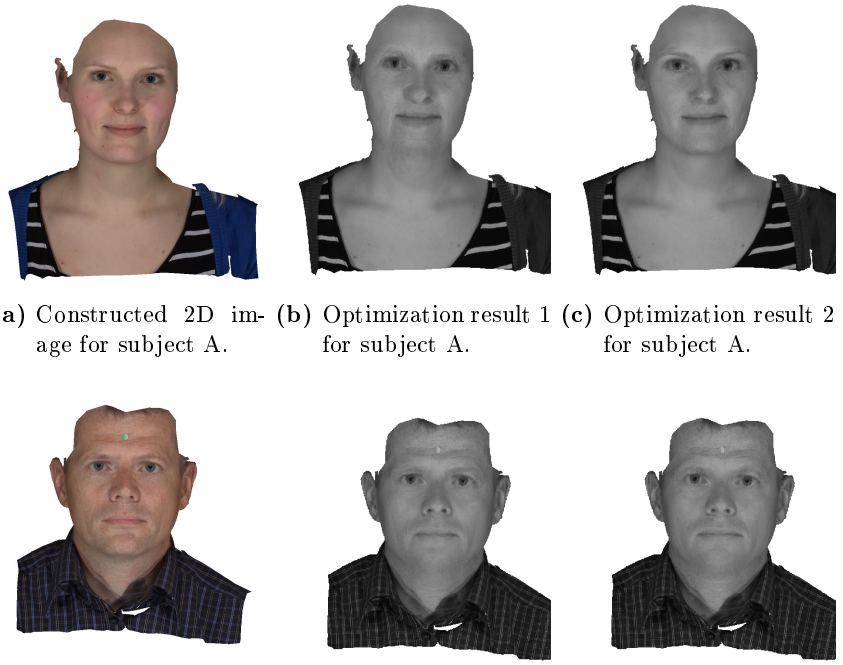
$$E_I = \sum_{x,y} ||I_{input}(x,y) - I_{model}(x,y)||^2 \tag{5.5}$$

is minimized, where $I_{model}(x,y)$ is the 3DMM projected into 2D. The method can be extended into including multiple 2D images as explained in [VRSV11].

The methods proposed by Blanz, Vetter and Paysan [BV03, PKA+09] aim to estimate the 3D shape and texture, along with the 3D pose and lighting conditions from a single photograph by fitting the 3DMM. The model fitting is only based on pixel intensities. Romdhani and Vetter [RV05] proposed a method that includes other image features (edges, highlights etc.) in the estimation of shape and texture. Our approach is to start by using the basic method using only texture and shape [BV03] and extend it if necessary.

In practice, fitting a 3DMM to a 2D image turned out to be a difficult task. This section shows our results when we simplify the problem by constructing an artificial '2D image'. The artificial 2D image is constructed by projecting a 3D face scan into 2D. Two such constructed 2D images are seen in Figure 5.3a and d. This procedure assures that the obtained 2D image is frontal, and the camera parameters are therefore known. Furthermore, because the 2D image is constructed from a 3D scan, we can assume that the lighting conditions in the 2D image correspond to those in the 3DMM. Even though the used 3D scans are captured by varying light conditions we assume that lighting conditions can also be neglected, which simplifies the optimization problem even further.

We optimize the weighting parameters α and β by minimizing Equation 5.5. Our optimization is performed using a standard unconstrained quasi-newton optimizer. By providing random start parameters the stability was tested and the experiments revealed that the optimizer is prone to finding local mimima. Figure 5.3 shows results for two individuals, referred to as subject A and subject B. Two optimization results are shown for each subject. The optimization results for Subject A vary a lot depending on the start parameters, in particularly in the nose region which can be observed in the figure. The results obtained for subject B are on the other hand very stable.



(a) Constructed 2D image for subject A. **(b)** Optimization result 1 for subject A. **(c)** Optimization result 2 for subject A.

(d) Constructed 2D image for subject B. **(e)** Optimization result 1 for subject B. **(f)** Optimization result 2 for subject B.

Figure 5.3: Results of fitting a 3DMM to a 3D scan projected into 2D.

5.4 Final remarks and future work

We generate a 3DMM based on approximately 600 face scans of individuals from the Danish Blood Donor Study. Our model does therefore include more variation than the publicly available Basel Face Model, which is based on 200 individuals. We showed that the largest modes of variation for shape and texture were related to height/width variations of the face and to light/dark variations, respectively.

Our work on fitting our 3DMM to 2D images is preliminary. We showed that even when we exclude camera parameters and lighting conditions our optimization scheme is still very unstable. Future work includes alternative optimization methods, using other image features and light modeling.

The methods and work just described was developed for the facial region, how-

ever we are interested in estimating full 3D head models based on 2D images. The next section describes our preliminary work on building a 3DMM for the head. As explained, the most complicated task when building a 3DMM is to create a dense point correspondence among the set of geometries, in particular for the complicated geometry of the pinna. In the following, we are therefore focused on creating correspondence between 3D pinna geometries.

5.5 A statistical outer ear model

We are interested in building a statistical 3D ear model to enable an estimation of 3D ear shape based on 2D images. A statistical 3D ear model is however also interesting for purposes outside the objectives of this thesis, for example biometrics and computer graphics.

The field of biometrics deals with recognition and identification based on different biometrics as for example a fingerprint or face images. It has been shown that the outer ear can be used for recognition purposes [BB02]. Most of the work on ear biometrics uses 2D images, however [CB07], [CB05] and [YB07] uses 3D range data along with 2D images. Yan et al. [YB07] uses an ICP based approach for matching 3D range images. Chang et al. [CBSV03] and Victor et al.[VBS02] perform a PCA on 2D ear images to generate a statistical ear model for recognition, referred to as 'eigen-ears'. The two papers show different results regarding recognition performance of the 2D eigen-ears. It would be interesting to combine 3D information with the statistical approach to generate 3D eigen-ears.

Computer graphics are already generating plausible human faces and body poses, for use in e.g. computer games, using 3D statistical models. Figure 5.4 shows the *facegen* model [fac]. The model shows different modes of faces, but the ears remain unchanged. To the best of our knowledge, no statistical ear models based on real 3D scans of humans exist.

5.5.1 Building our statistical 3D ear model

The approach used in Section 5.2 for building a statistical face model, is also applicable to ears. Instead of 3D scans of faces we use 3D ear models, which have been cut out of the 3D head models from the DTU Compute 3D head database, Section 4.5. We build our statistical 3D ear model based on the 11 ears shown

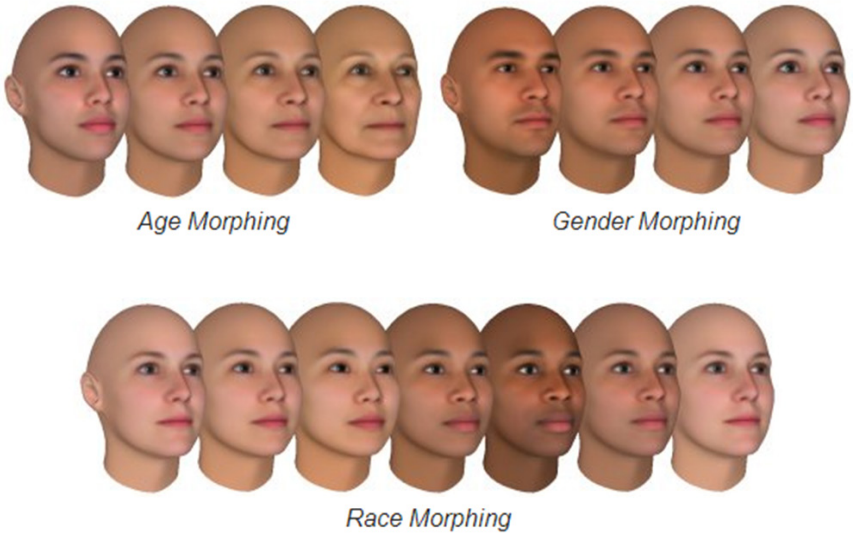


Figure 5.4: Screenshot from [fac].

in Figure 5.5. A large shape variation can be seen between the different ears for example in the attachment of the ear lobe or in the 'straightness' of the ear.

Our main obstacle is to create a dense point correspondence between the 3D ear models. It is straightforward to create correspondence between simple, non bending, surfaces using a minimum distance approach and a few manually annotated landmarks [PLN+02]. However, as seen in Figure 5.5, the outer ear is a very complex structure with several bendings of the surface. Such deep bendings of the surface will cause a minimum distance approach to fail.

Our approach is, therefore, to use local shape features to create a sparse correspondence which can guide a non-rigid registration of the ears. Ones the ears are registered a dense point correspondence is easily obtained by projecting the vertices from one ear to another. We utilize a Point Feature Histogram (PFH) as our local shape feature, because the PFH is a simple descriptor, which describes local 3D geometries of the ear well. The PFH that we employ is a Point Cloud Library (PCL) implementation [RC11].

Figure 5.5: Ear shapes used for building the statistical 3D ear model.

5.5.1.1 The Point Feature Histogram

The Point Feature Histogram (PFH), developed by Rusu et al. [RMBB08], provides a simple description of the geometry of the k-neighborhood of a point. The PFH describes the neighborhood using only a few parameters, which makes the descriptor fast to calculate.

The PFH for two points p_s and p_t with corresponding normals n_s and n_t is based on four features, that measure the angle difference between the point normals and the distance vector between them. Initially a local fixed coordinate system is defined as [RMBB08]:

$$u = n_s \tag{5.6}$$
$$v = (p_t - p_s) \times u \tag{5.7}$$
$$w = u \times v. \tag{5.8}$$

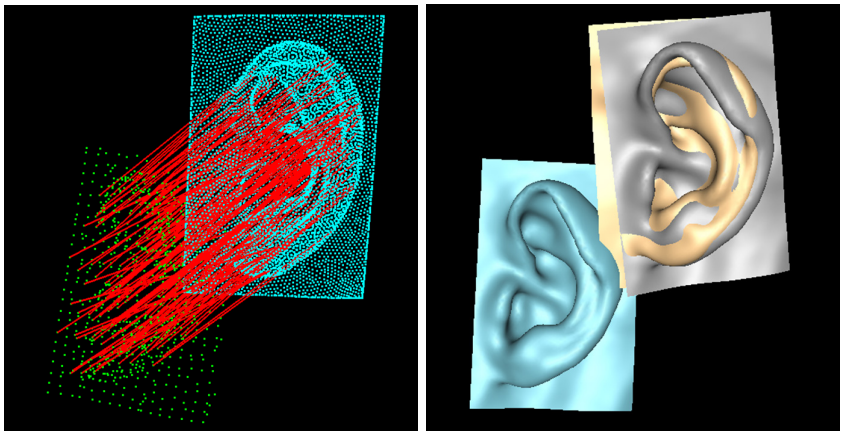
Given the (u, v, w) coordinate system the four features are given by [RMBB08]:

$$f_1 = \langle v, n_t \rangle \tag{5.9}$$
$$f_2 = ||p_t - p_s|| \tag{5.10}$$
$$f_3 = \frac{\langle u, p_t - p_s \rangle}{f_2} \tag{5.11}$$
$$f_4 = \text{atan}(\langle w, n_t \rangle, \langle u, n_t \rangle), \tag{5.12}$$

where $\langle ., . \rangle$ describes the dot product. The four features are binned into a descriptive histogram. In the PCL implementation the f_2 feature is omitted.

A sparse set of correspondences are created between the two ears that should be registered. Figure 5.6a shows correspondences between two ears, or point clouds. The green point cloud corresponds to the blue ear in Figure 5.6b and will be referred to as Ear1, whereas the blue point cloud corresponds to the skin colored ear and will be referred to as Ear2. We seek to register Ear1 to Ear2. Correspondences are created based on the minimum Euclidian distance between PFHs, and false correspondences are removed based on a random sample consensus (RANSAC) [FB81] rejection. An initial rough alignment is obtained by an affine transformation, which is calculated based on the sparse correspondences. The gray ear in Figure 5.6 is Ear1 after the affine transformation.

Now that Ear1 has been roughly aligned to Ear2, Ear1 is non-linearly warped to Ear2 using a TPS warp as described in [PLN+02]. Figure 5.7b shows Ear2 and a TPS warped Ear1. The TPS warp is also guided by the set of sparse correspondences. Because the TPS warp has changed the shape of Ear1 a new set of sparse correspondences is calculated. The new set forms the basis for



(a) Correspondences between Ear1 (green points) and Ear2 (blue points) found as the minimum Euclidian distance between PFHs.

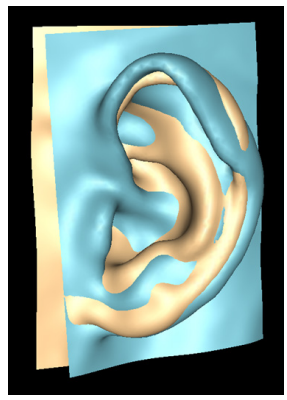
(b) Affine transformation of Ear1 (blue) to Ear2 (skin) based on the correspondences seen in a). The gray ear is the transformed Ear1.

Figure 5.6: Initial rough alignment of ear geometries based on an affine transformation.

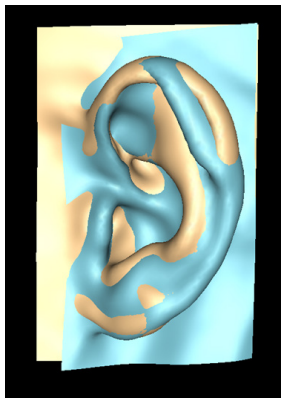
another TPS warp and the process is iterated six times, because six iterations seemed adequate for our purpose. The results from each iteration is seen in Figure 5.7. After the final warp a dense point correspondence is established by projecting vertices from Ear1 onto Ear2.

The procedure is performed for all 11 ears in Figure 5.5 and a Procrustes alignment and PCA are applied to obtain a statistical model. The three largest modes of variation are shown in Figure 5.8. The behavior of the PCs is difficult to interpret, since every PC appear to include more than one aspect. The first PC appears to explain the straightness of particularly the upper part of pinna. The second PC seems to account for the width of pinna, whereas the third PC is difficult to interpret.

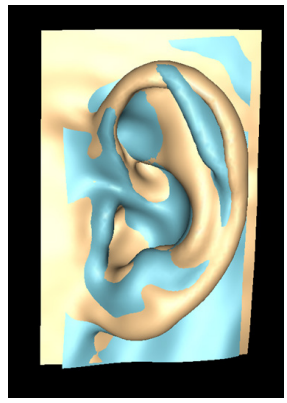
The behavior of a statistical 3D ear model is interesting in itself, however our goal was to create a statistical 3D head and ear model and to fit the model to one or several 2D images to create individual head and ear models. This problem remains to be solved.



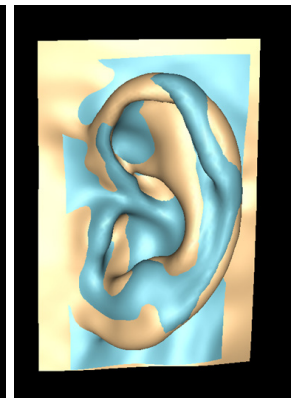
(a) Affine transformed.



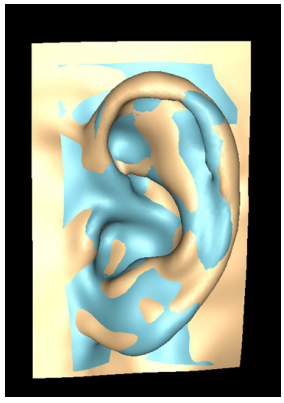
(b) Iteration 1.



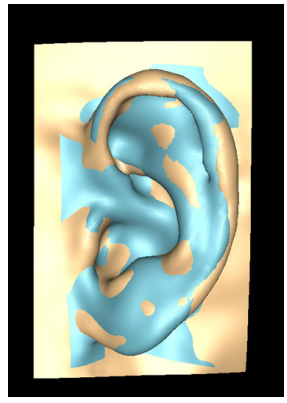
(c) Iteration 2.



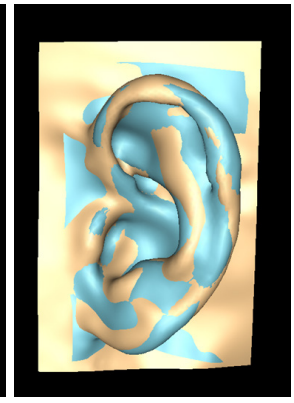
(d) Iteration 3.



(e) Iteration 4.

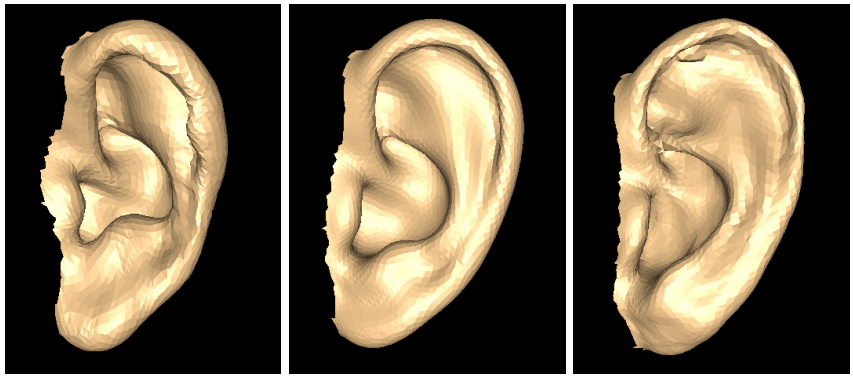


(f) Iteration 5.

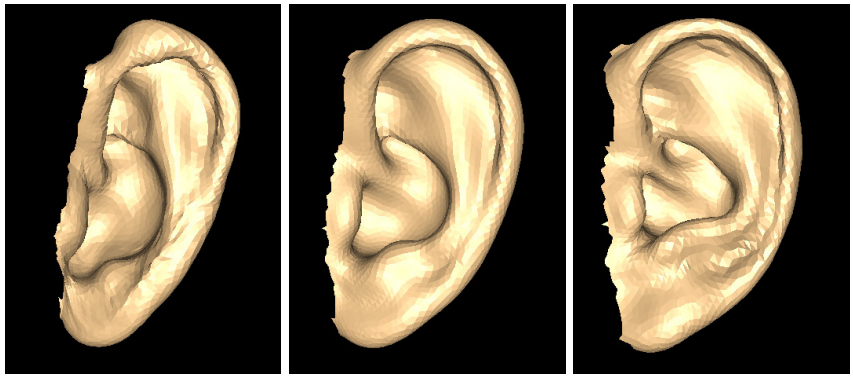


(g) Final transformation.

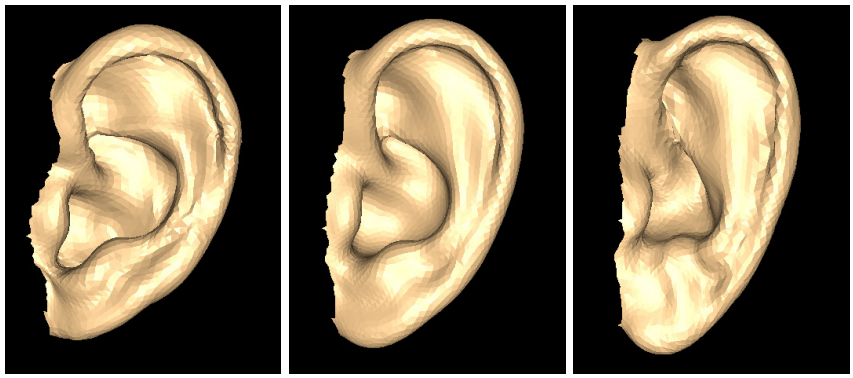
Figure 5.7: Iterative TPS warp of Ear1 (blue) to Ear2 (skin). The warp is guided by a set of sparse correspondences found using PFHs.



(a) First PC.



(b) Second PC.



(c) Third PC.

Figure 5.8: The three PCs with the largest shape variation. Center image: mean ear, Left image: mean ear minus 3 SDs, Right image: mean ear plus 3 SDs. Note that these ears are cut, which was not the case earlier.

5.5.2 Final remarks and future work

We built a statistical model for the left outer ear based on 11 ear models. In order for the statistical ear model to explain the variance of the population the model should be extended to consist of several hundred ear models. We could with time extend the model by scanning more individuals.

The method that we developed to create a dense point correspondence between two ear geometries is not very robust, and it fails if the ear shapes are too different. The most common problem for the algorithm is to create correspondence between ears with different earlobes (free vs. attached earlobe). The reason is, that the PFHs for a free and an attached earlobe are very different, and that the earlobe regions are therefore not matched. The problem could be solved by generating a statistical distribution of PFHs for the earlobe region, and to sample from this distribution instead of using the PFHs calculated for the template in this particular area.

Much work remains before we are able to estimate 3D head models based on 2D images. The statistical ear model should be extended to include the entire head area and to include a few hundred individuals. Furthermore, the process of fitting a 3D model to 2D images should be extended to complete head models. Future work could also include an investigation of other image modalities instead of 2D images. It would for instance be interesting to investigate the use of depth data obtained by a Kinect, because the Kinect is easily accessible and inexpensive enough to have in a clinic. In the future smartphone cameras might also be used to generate a rough initial surface scan.

Chapter 6

A 3D printed listener-specific head simulator

This chapter presents the work that has been performed on 3D printing (more formally known as additive manufacturing, see Section 6.2) a real size listener-specific head simulator. The work comprise computer-aided design (CAD) modifications of the 3D head model obtained in Chapter 4, 3D printing of the head and ear parts, and an evaluation of the final printed head.

The CAD modifications was conducted by Peter Frederiksen from Oticon A/S and Rasmus R. Paulsen. The 3D printing of the head part was executed at the DTU Fablab [fab] and the 3D printing of the ear parts was performed by Damvig Develop A/S. The planning of the printing process, material properties and other requirements was a collaboration between the manufactures, Martin Larsen, Rasmus R. Paulsen and the author. Measurements performed on the printed head for evaluation purposes was conducted by Jonas L. Olesen at Oticon A/S.

6.1 Generic head and torso simulators

HRTFs are often measured on head and torso simulators [GM+94, RBW95], because of the very cumbersome and time consuming measurements that are required to obtain listener-specific HRTFs [MBL07]. A listener-specific HRTF is a HRTF measured for the individual human subject. Two very popular head and torso simulators are the KEMAR produced by G.R.A.S. Sound and Vibration [KEM] and the HATS produced by Brüel & Kjær [HAT]. Both head and torso simulators are produced to provide a realistic reproduction of the acoustic properties of an average adult human. The KEMAR is furthermore build on a large statistical research of the average human body. Both head and torso simulators are seen in Figure 6.1.

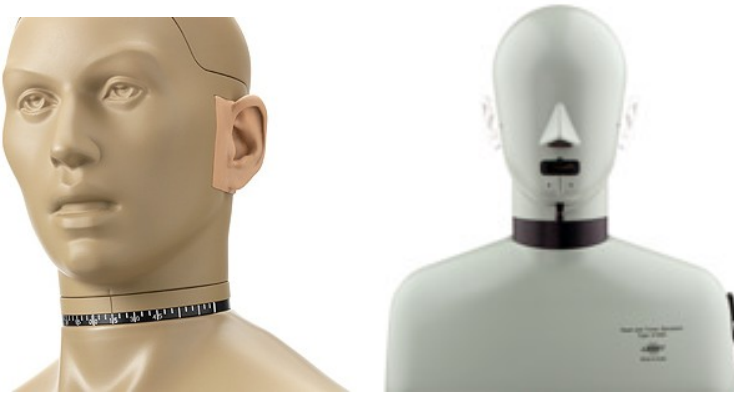


Figure 6.1: Left: KEMAR made by G.R.A.S., Right: HATS made by Brüel & Kjær. From[1]

Even though both KEMAR and HATS represent an average human adult, neither of them have the shape of an actual human subject as a listener-specific head simulator would have. A listener-specific head simulator is interesting for several purposes, especially within hearing aid or headset development and research. It facilitates testing of different hearing aid or headset features on a real human geometry. The directional microphone is for instance a hearing-aid feature which is dependent on the individual head and ear shape. Furthermore, a small collection of individual 3D heads would provide the basis for optimizing directionality, or any other feature, for different characteristic groups of individuals (men, women, children).

[1] `http://www.gras.dk/45bb.html` and `http://www.bksv.com/Products/transducers/ear-simulators/head-and-torso/hats-type-4128c`

With respect to this thesis, the purpose of the listener-specific head simulator is to enable an extra verification of the directionality obtained from simulations, see Figure 1.2. Comparing results obtained from measurements on the printed head with simulations based on the same head model removes some of the differences that would exist if human measurements were used. Geometrical differences are reduced because the printed head and simulations are based on the same head model and on the same torso simulator. Human acoustical properties are replaced with properties of the printing material, which can be assumed to be more similar to the properties of the simulated model. Furthermore, human movements, caused by for instance breathing are excluded. All of which results in a more righteous comparison of measurements versus simulation.

With the rapid development in additive manufacturing new possibilities are emerging. Previously, creating an accurate replica of a human head would have required massive amounts of work. It is still not a trivial task, but with the development of software tools and the decreasing costs of 3D printing, generating flexible human replicas for product testing is becoming increasingly feasible.

6.2 Additive manufacturing

Additive manufacturing (AM), formerly known as rapid prototyping, is a technology where a CAD model is produced directly, without the need for process planning. The name additive manufacturing refers to the basic principle of AM, that the model is build by adding a layer at a time. Opposite, traditional manufacturing processes are often subtractive instead of additive. AM is therefore not only a fast method, it is also material-saving. AM is less formally called 3D printing, and the two terms will be used indiscriminately in the reminder of this thesis. The AM theory presented here is described in various texts, however this section is mainly based on [GRS+10].

3D printing a model requires, in general, eight steps:

1. The closed surface geometry of the model is fully described, for instance using a CAD software.
2. The CAD model is converted to the file format stereolithography (STL).
3. The STL file is transferred to the AM machine, and the STL file is adjusted for size, position and orientation.
4. The AM machine is setup with respect to material constraints, power of the energy source, layer thickness etc.

5. The model is automatically build.

6. The printed part is removed from the machine.

7. The part is post-processed depending on the printing method (supporting structures are for instance cleaned away).

8. The part is ready for application.

The AM process described above is a general approach. The machine setup, build, and post-processing are highly dependent on which AM machine or process that is used. AM processes can be divided into seven categories [GRS+10]: vat photopolymerization, powder bed fusion, material extrusion, material jetting, binder jetting, sheet lamination and directed energy deposition. The general principle being that material is deposited in layers and solidified in some way. The different methods handle different types of material such as polymers, metals and ceramics. Due to our requirements for material properties, cost and needed accuracy we chose Selective Laser Sintering (SLS) printing, which is a powder based fusion method, for the manufacturing of our listener-specific head simulator.

6.3 Our listener-specific head simulator

We produce a listener-specific head simulator using 3D printing. The 3D printed head is based on one of the 3D surface models obtained in Chapter 4. More specifically, we used the head model in the upper left corner of Figure 4.6, referred to as subject NH167. The surface model requires some modifications before it can be used for 3D printing or for the acoustical modeling presented in Chapter 7. The modifications and the 3D printing will be explained in the next two sections. The complete process of producing the listener-specific head simulator is also presented in Paper B.

6.3.1 CAD modifications

The original surface model, seen in Figure 6.2 left, consists of vertices and triangles. The surface model should in addition to the 3D printing, also be used for the numerical simulations in Chapter 7. Several numerical simulation packages exist that can use triangulated surfaces as input but most

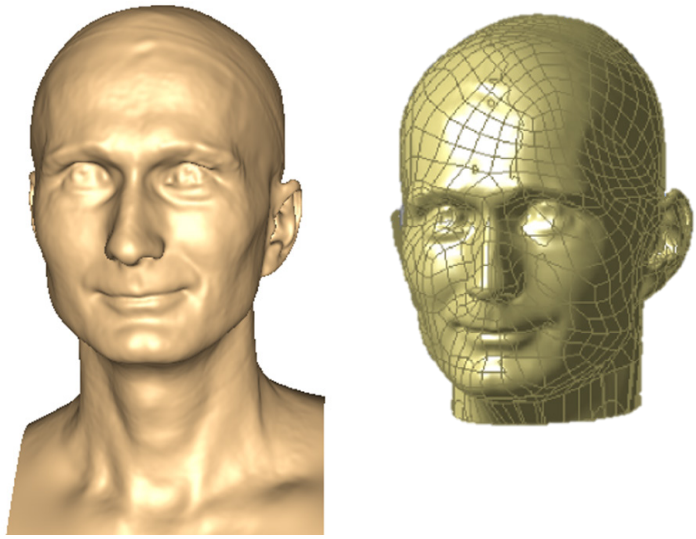


Figure 6.2: Left: surface model of test subject NH167, Right: CAD modified model ready for printing.

programs work better with surfaces represented as non-uniform rational B-splines (NURBS) [FVDF+94]. The 3D surface model was therefore converted to NURBS [NUR].

The current surface representation of the head model has no thickness, which is required for the printing process. An overall minimum thickness of 5 mm was added to the surface model. The thickness was added by creating a balloon-like object inside the head, which was manually reshaped to turn the head into a hollow shell. Certain parts like the nose remained solid.

In order to mount a torso simulator (Type 4128, Brüel & Kjær) on the head after the printing, a flange was added to the model at the bottom of the head. Furthermore, in order to enable separate manufacturing of head and ears using different materials, boxes around the ears were created. After manufacturing, the ears were attached to box-shaped elements that were inserted in the corresponding box-shaped recesses in the head. The final head model ready for printing is seen in Figure 6.2 right.

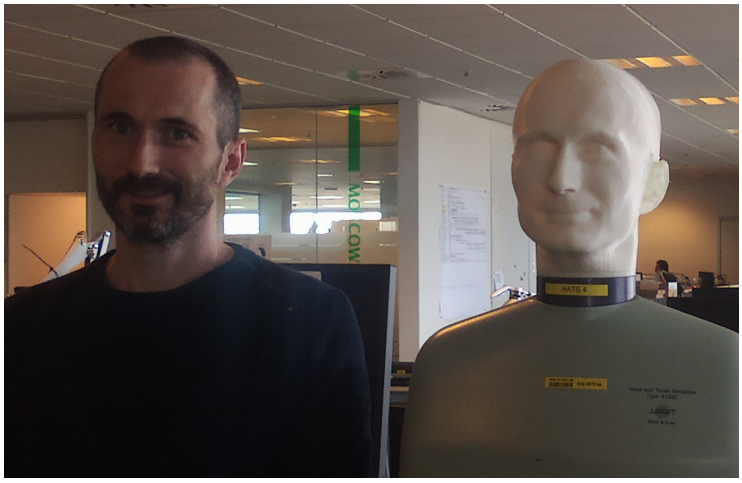


Figure 6.3: Test subject NH167 with the listener-specific head simulator based on subject NH167's head model.

6.3.2 3D printing

In the current work, we have not focused on creating a full head with material properties that exactly match a living human. It would be very complicated and it is believed that the impact on the acoustical measurements is of less significance than other factors in the measurement setup. The head was therefore printed in hard plastic except the ears, which were printed in a soft material. The reason for choosing a soft material for the ears was mainly the need of placing hearing aids behind the ears in a way that simulates the position of hearing aids on a human subject.

The head part: The head part was printed at DTU FabLab [fab]. The hard part of the head was printed on an SLS printer EOS P395 [eos] in hard acrylic (PA2200) using 0.12 mm layers. The head part, with ears and torso mounted, is seen in Figure 6.3. The figure also shows test subject NH167, who the printed 3D head model was based on, for comparison.

The ears: The ears were printed at Damvig Develop A/S. The ears used for our acoustical measurements in Chapter 8 were printed on a polyjet Objet500 Connex 3D printer [str] using two materials and a layer thickness of 0.03 mm. The core was printed in a hard, acrylic material (Verowhite - FullCure830) and the outer part was printed in a soft material (TangoPlus - FullCure930). The ears are seen in Figure 6.4 left.

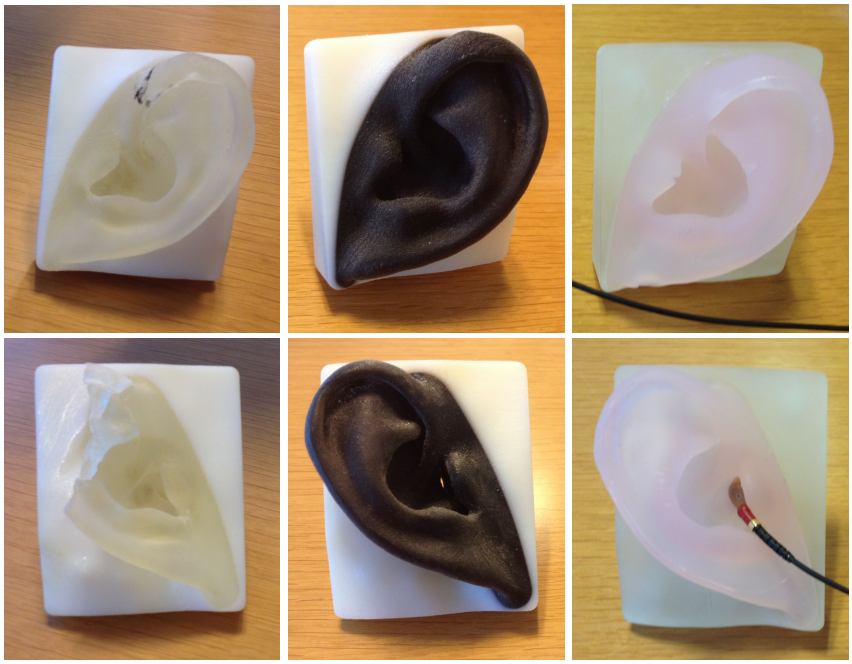


Figure 6.4: Three sets of ears manufactured for the listener-specific head simulator.

Because the first set of ears had a tendency to disintegrate, two additional sets of ears were manufactured. The second set of ears were printed, whereas the third set of ears were manufactured by a combination of 3D printing and molding. Here, the core was printed in a hard acrylic material and the pinna was molded in silicone (Egger Flex /AB - silicone shore 25, with Visijet sl impact fixture), to achieve softer material properties. The two additional sets of ears are seen in Figure 6.4 center and right.

6.4 Evaluation of the printed geometry

The 3D listener-specific head simulator is subject to errors arising both from the scanning, stitching and meshing method described in Chapter 4, and from the CAD modification and printing process described in this chapter. The accuracy of the head model was, with the best of our abilities, evaluated in Section 4.4. It was however not easy to evaluate a human head model because of a lack of known head geometry (golden standard). Now that the human 3D head model has been

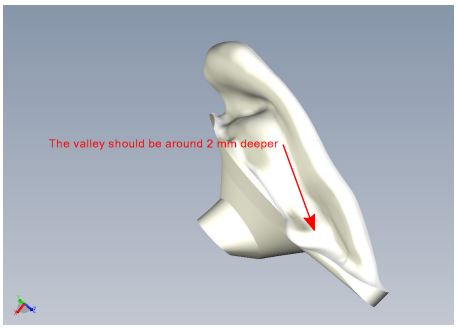
printed into a physical model, a golden standard or correct geometry exists in the form of our human test subject. It is not an easy task to compare a printed head with a human head, however measurements between certain fix points on the two geometries can give an idea of the quality of the printed head. The pinna regions are of greatest concern with respect to generating individual directionality, since pinna is believed to assist in sound source localization [Bla83]. Our evaluation therefore has a focus on evaluating the geometry of the left and right pinna.

The first set of printed pinnas were compared with the actual human pinnas by measuring distances with a vernier caliper. Four problem areas, indicated by red arrows in Figure 6.5, were located on the two pinna geometries:

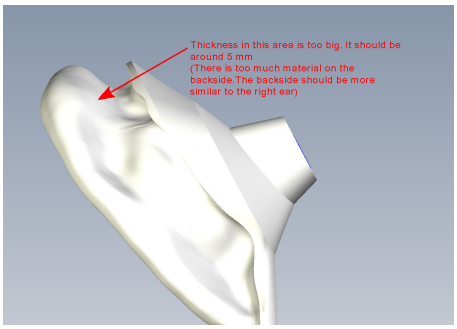
1. **Right ear:** The valley should be around 2 mm deeper.
2. **Left ear, error 1:** Thickness in this area is too big. It should be around 5 mm (There is too much material on the backside. The backside should be more similar to the right ear).
3. **Left ear, error 2:** The valley should be around 2-3 mm deeper.
4. **Left ear, error 3:** Around 1-2 mm should be removed in the curve (depth) between the two points.

The errors are mostly located on the back side of the pinna, and the errors are in the 1-3 mm range. Errors smaller than 1 mm are neglected because they might arise from vernier caliper measurement errors. The right pinna is in general more accurate than the left pinna.

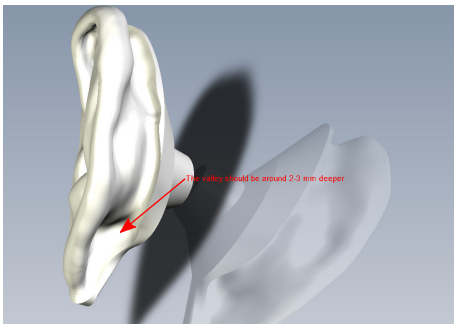
The errors presumably arise from the surface reconstruction approach described in Chapter 4. As explained the pinna geometry is difficult to scan using a surface scanner, because the complicated geometry leads to many occluded areas. The occlusions result in areas that are not well covered with data and the reconstruction approach are forced to interpolate in these areas. The evaluation of the pinna geometries in Figure 6.5 suggests that the areas behind the pinna are in particularly difficult to scan, leading to the measured errors. The same problem was noted in Section 4.4 where the model was evaluated against a head model of KEMAR.



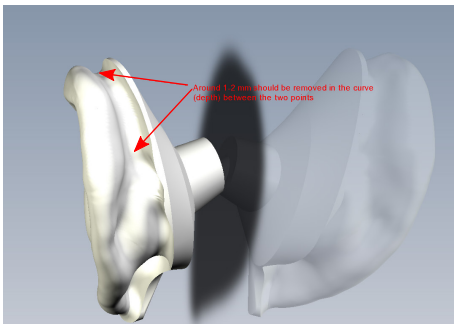
(a) Right ear.



(b) Left ear, error 1.



(c) Left ear, error 2.



(d) Left ear, error 3.

Figure 6.5: Evaluation of ear geometries. Because the comments are unreadable in the figures, they are stated as bullet points in the text, named after the subplot titles.

Chapter 7

Acoustical modeling

This chapter contains a theoretical part on acoustical modeling, with a focus on the finite element method (FEM) and a methodological part that explains how we have employed FEM for acoustical modeling of HRTFs and for structural modeling of BTE hearing aid placement. The theoretical part is explained using the notation from Dhatt et al. [DLT12].

The work described in this chapter contains three different FEM simulations: a static structural simulation for BTE placement, a harmonical simulation for the BTE microphones and a harmonical simulation for microphones placed in the ear canal. All simulations were conducted using Ansys. The static structural simulation was setup and conducted by Martin Larsen, whereas the two harmonical simulations were a collaboration between Martin Larsen and the author.

Paper B and Paper D accounts for the different methods and results obtained throughout this chapter.

7.1 Acoustical modeling of HRTFs

HRTFs are, as explained in Chapter 1, very cumbersome and time consuming to measure. It is therefore of great interest to be able to simulate and model HRTFs in silico. The major goal is to simulate HRTFs based on 3D geometries of individual human heads. Different methods exist for acoustical modeling of HRTFs. Anthropometric measures [ZHDD03], Principal Component Analysis [KW92] and pole-zero modeling [Jen95] are all methods that have been used for prediction of HRTFs. All of these methods use a database of HRTFs to extrapolate either an entire set of HRTFs or HRTFs for missing source positions. The required database is not always available and other methods than extrapolation are needed. Furthermore, using only simple 2D or 3D anthropometric measures seems insufficient, since it is known that the complex 3D shape of pinna has a large influence on the HRTF's structures.

The shape of the entire head, pinna, and torso can be used for modeling HRTFs using either the boundary element method (BEM) or finite element method (FEM). BEM has been used for calculating individual HRTFs [Kat01, OHI09, HVH+14] and the method seems to be the natural choice for acoustical modeling, since BEM calculations are often faster than FEM. However, we are not only interested in acoustical modeling, since we also want to calculate the correct bending of pinna when a BTE hearing aid is placed on top of the ear. The BTE placement is a structural analysis, which cannot be modeled using only elements on the boundary, as it is the case in BEM. We therefore employ FEM for all of the simulation work performed in this thesis.

The next section contains a theoretical explanations of FEM simulations. However, because we employ FEM without developing any new methods this theoretical part can be skipped if the reader is familiar with FEM, and the reader can jump to Section 7.3.

7.2 The Finite Element Method

Any physical system can be modeled using partial differential equations, and they can be solved numerically. Numerical modeling of a system usually involves four steps:

- A description of the physical system using engineering terms.
- Translation of the system into a mathematical model (partial differential

equations).

- Construction of a numerical discrete model that can be solved using a computer (algebraic equations).
- Computer code development to simulate the behavior of the system.

FEM is employed when the set of partial differential equations must be discretized and transformed into algebraic equations. The principle of FEM is to use a simple, for instance linear, approximation of unknown variables to transform the partial differential equations into algebraic equations, that are easily solved using a computer. Because we use the commercial program ANSYS for our simulations, we define our physical system on the software's interface and the remaining steps are handled by the program. It is however also important to choose an appropriate discretization, which we will return to in Section 7.3.

7.2.1 Approximation with finite elements

A physical system can be described using a mathematical model with a number of variables $u_{ex}(x)$. The variables can be represented by 'approximate' functions $u(x)$. The difference between the approximate functions and the actual variables, $e(x) = u(x) - u_{ex}(x)$, should be small enough to satisfy the purpose of the desired problem. In our simulations we want to have a mesh that is fine enough to represent our surface model accurately. Furthermore the acoustical simulations have requirements on the number of elements per wavelength, see Section 7.3.

The approximation function $u(x)$ can profitably be constructed using finite elements. Figure 7.1 shows a one dimensional approximation case using four nodes, and a linear interpolation between the nodes. In FEM the domain V is separated, using nodes, into subdomains, or elements, V^e. Different approximate functions $u^e(x)$ are constructed for each element. The nodal approximation over an element V^e is constructed so that it only depends on the nodal variables placed within or on the boundary of the given element. Furthermore, the approximate function $u^e(x)$ should be continuous over the element V^e and satisfy conditions of continuity between adjacent elements.

The approximation function $u(x)$ is constructed from a set of basis functions N as [DLT12]:

$$u(x) = N_1(x)u_1 + N_2(x)u_2 + ... + N_n(x)u_n \tag{7.1}$$

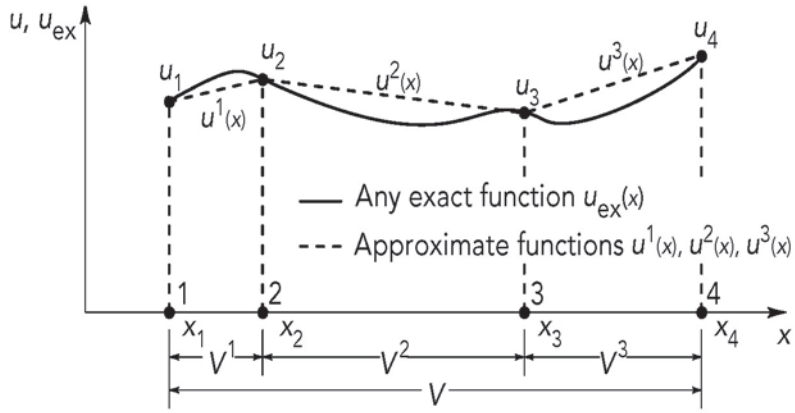


Figure 7.1: One dimensional finite element approximation. From [DLT12]

or in matrix notation as:

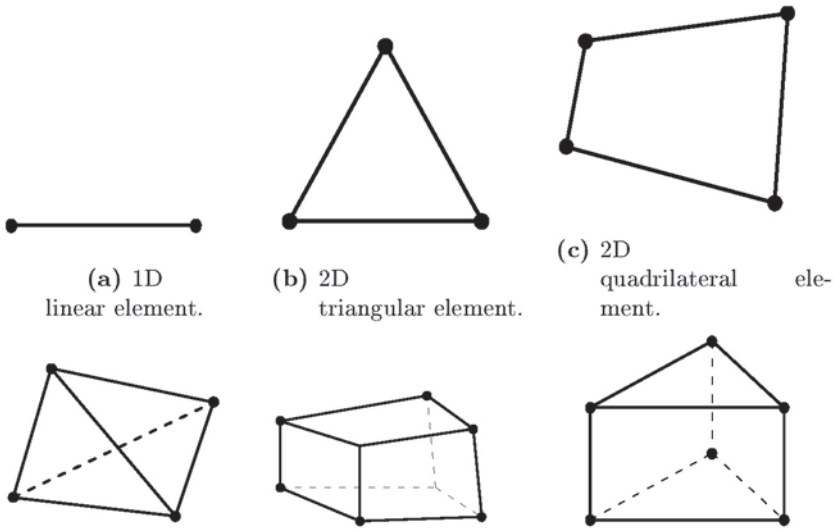
$$u(x) = \langle N_1(x)\ N_2(x)\ \ldots\ N_n(x) \rangle \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{Bmatrix} = \langle N(x) \rangle \{u_n\} \tag{7.2}$$

In Figure 7.1 the approximation function $u^1(x) = N_1 u_1 + N_2 u_2$, where $N_1 = \frac{x-x_2}{x_1-x_2}$ and $N_2 = \frac{x-x_1}{x_2-x_1}$ is a simple linear interpolation for $x_1 < x < x_2$.

7.2.1.1 Elements

When a domain V is partitioned into elements V^e, they must satisfy two rules [DLT12]:

- Two elements can only have common points on their common boundaries, if such boundaries exist (No overlapping).
- The set of elements V^e must constitute a domain that is as close as possible to the original domain V (the domain should be filled). There should be no 'holes' between elements.



(a) 1D linear element.

(b) 2D triangular element.

(c) 2D quadrilateral element.

(d) 3D tetrahedral element.

(e) 3D hexahedral element.

(f) 3D prismatic element.

Figure 7.2: Linear elements in 1D, 2D and 3D. From [DLT12].

Many different elements exist with different properties. Figure 7.2 shows 1D, 2D and 3D linear elements, but elements can also be quadratic or cubic. Where a linear element only has nodes at the corners of an element, a quadratic element also contains midside nodes and a cubic element contains two additional nodes on each element side. The number of nodes are proportional with the number of equations that need to be solved, and the linear elements are therefore less computationally heavy. We use linear elements for the acoustical calculations, however cubic elements are required for for the mechanical calculations, because linear elements perform poorly for mechanical problems.

When a physical problem is solved using FEM a mesh is generated for the entire domain V. Because each element has individual approximate functions, and thereby individual element shapes, calculations must be performed for each element separately. This is computationally expensive, especially because the element shape might be complex. To reduce computational costs, and to simplify the analytical description of an element, a reference element is introduced. The reference element is an element with a very simple shape, Figure 7.3 shows a reference element in the case of a triangle.

A reference element, V^r can be transformed into an element V^e by a geometrical transformation τ^e:

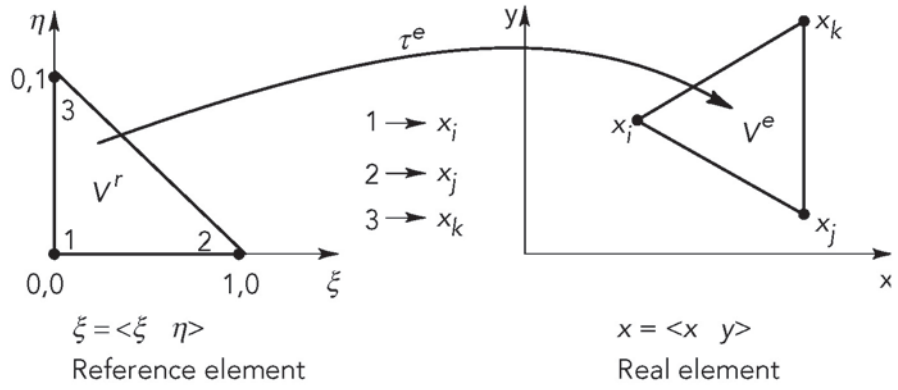


Figure 7.3: Transformation from reference element to real element. From [DLT12].

$$\tau : \xi \rightarrow x(\xi) = \langle \overline{N}(\xi) \rangle \{x_n\}, \tag{7.3}$$

where $\overline{N}$ is called a geometrical transformation function.

In the case of Figure 7.3 the transformation becomes:

$$x(\xi, \eta) = \overline{N}_1(\xi, \eta)x_i + \overline{N}_2(\xi, \eta)x_j + \overline{N}_3(\xi, \eta)x_k \tag{7.4}$$

$$y(\xi, \eta) = \overline{N}_1(\xi, \eta)y_i + \overline{N}_2(\xi, \eta)y_j + \overline{N}_3(\xi, \eta)y_k. \tag{7.5}$$

The approximation function from Equation 7.2 can now be expressed using reference elements:

$$u_{ex}(\xi) \approx u(\xi) = \langle N(\xi) \rangle \{u_n\}. \tag{7.6}$$

Notice that a reference element, for instance a triangle, can be applied to all elements of that particular shape. We use a mesh consisting of hexahedral and tetrahedral elements and since we employ two different kinds of elements only two reference elements are required.

7.2.2 Matrix formulation of FEM

FEM use an element wise approximation of a system to discretize and transform differential equations into algebraic equations. The organization of the

elementary equations into matrices are important, since matrices are efficiently employed in computer codes.

In FEM a description of a physical system using differential equations are referred to as *strong*, whereas an expression of the system using an integral formulation is referred to as *weak*. The weighted residual method is used to generate weak, integral formulations and the Galerkin method solves the (weak) problem using discretization.

In general the weak expression can be formulated as:

$$W = \int \int_A ...dA + \int_{S_u} \delta u.f_R ds + \int_{S_f} \delta u.f_{ext} ds, \tag{7.7}$$

where A is the domain of calculation, and S_u and S_f are parts of the surface, where external displacements and surface loads are applied.

Using the approximation with finite elements, which was introduced in Section 7.2.1, the domain of the calculation becomes:

$$V = \sum_e V^e \quad \text{and} \quad x^e = \sum_i \overline{N}_i(\xi, \eta, ...)x_i, \tag{7.8}$$

where ξ and η are elementary coordinates in the reference space, x^e is a vector of elementary coordinates, x_i is a vector of elementary nodal coordinates and $\overline{N}_i$ are geometric interpolation functions. The domain V is therefore expressed as a sum over elementary domains, and the coordinates of an element is expressed as a sum over nodal coordinates.

The approximate solution and test function for each element is defined as:

$$u^e(x,t) = \sum_i N_i(\xi, \eta, ...)u_i(t) \tag{7.9}$$

$$\delta u^e(x) = \sum_i N_i(\xi, \eta, ...)\delta u_i, \tag{7.10}$$

where $u_i(t)$ and δu_i are the nodal variables of the solution and test function, respectively.

The elementary weak form from Equation 7.7 can now be discretized using finite elements. For a structural dynamic problem:

$$W_h^e = \langle \delta u \rangle \left([m] \left\{ \frac{d^2 u_i}{dt^2} \right\} + [c] \left\{ \frac{du_i}{dt} \right\} + [k] \{u_i\} - \{f\} \right), \tag{7.11}$$

where $[m]$, $[c]$ and $[k]$ are the mass, damping and stiffness matrices for a given element, f is the load vector for the element and r is the residual vector for the element.

The local contributions from each element can be assembled into a global system of equations:

$$W_h = \sum_e W_h^e = \langle \delta U \rangle \left([M] \left\{ \frac{d^2 U}{dt^2} \right\} + [C] \left\{ \frac{dU}{dt} \right\} + [K] \{U\} - \{F\} \right) = \langle \delta U \rangle R, \tag{7.12}$$

where $[M]$, $[C]$ and $[K]$ are the global mass, damping and stiffness matrices, F is the global load vector and R is the global residual vector. Furthermore, U is the global vector of unknown nodal functions. The solution of the problem is found when $R = 0$.

For a harmonical response analysis Equation 7.12 becomes [Koh13]:

$$W_h = \langle \delta U \rangle \left(\left(-\omega^2 [M_a] + j\omega [C_a] + [K_a] \right) \{p\} - \{F_a\} \right) = \langle \delta U \rangle R, \tag{7.13}$$

where $[M_a]$, $[C_a]$ and $[K_a]$ are global acoustic mass, damping and stiffness matrices, and $\{F_a\}$ is an applied acoustic force.

We use a static version of Equation 7.12 for our simulation of BTE hearing aid placement and Equation 7.13 for our harmonical simulations. We rely on the ANSYS solver to find a solution to the system of equations.

7.3 Our FEM simulations

We simulate HRTFs for two different microphone placements. The two simulation cases will be referred to as:

- **The microphone-ITE simulation:** A single microphone placed in each ear canal.
- **The BTE simulation:** Two microphones incorporated into a BTE shell positioned on the left ear.

The microphone placement is straightforward in the first simulation. However, in the second simulation a BTE hearing aid needs to be positioned. When a BTE hearing aid is positioned behind the ear in the physical world, the outer ear will deflect due to the presence of the hearing aid. We simulate BTE placement and corresponding pinna deflection in Section 7.3.1. Section 7.3.2 accounts for the harmonical response simulations for the microphone-ITE and BTE simulations.

All of our FEM simulations were setup in Ansys 15.0. Ansys is according to [DLT12] one of numerus state-of-the-art general-purpose computer codes that are available for industrial users. We choose Ansys for our FEM modeling due to the acoustics extension that is available.

7.3.1 BTE placement

The BTE placement was solved using a static structural simulation, because we assume that the displacements of the hearing aid are applied sufficiently slow. Equation 7.12 therefore becomes [DLT12]:

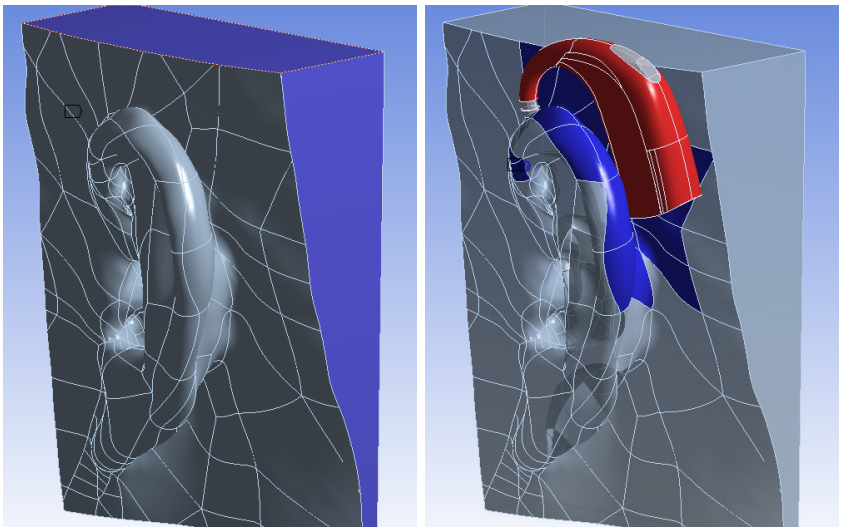
$$\{R\} = [K]\{U\} - \{F\} = 0, \tag{7.14}$$

where $[K]$ is a global stiffness matrix and $\{F\}$ is the applied displacement.

The position of a BTE hearing aid when worn is determined by its own shape and by the shape of the ear. In the physical world, the hearing aid is placed on the ear by pressing it down and letting go. The ear will deflect and hold the hearing aid in place, and the final position of the hearing aid will be determined by the frictional forces between the ear and the hearing aid.

The simulation attempts to imitate the physical world. A static structural analysis using an applied force to push the hearing aid into place was initially tested. The method did however not converge and a different approach was therefore employed. Instead of pushing the hearing aid into place using a force, a displacement was introduced to pull the hearing aid into place. After the displacement has been applied, the hearing aid is released and the stiffness and friction of the ear keeps the hearing aid into place.

The BTE placement was based on the 3D head model generated in Chapter 4 and modified in Chapter 6, and a 3D geometry of a standard BTE shell. The ear was cut from the rest of the head model to save calculation time. A mesh was generated for the simulation with 25,580 nodes for the ear and 4,754 nodes for the hearing aid. The ear model was constrained with fixed nodes on all the cut surfaces, see Figure 7.4a.



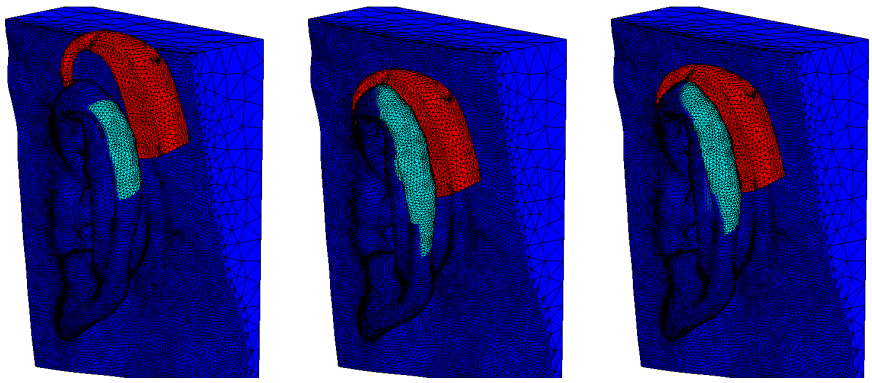
(a) Cut surfaces, which are fixed (blue faces).

(b) Contact surfaces (blue and red faces).

Figure 7.4: The BTE placement requires fixed surfaces and contact surfaces between ear and hearing aid to be defined. a) Shows the fixed surfaces in blue and b) shows the contact surfaces for the ear (blue) and hearing aid (red).

Our calculations consisted of three sub steps, see Figure 7.5. Figure 7.5a shows the initial position of the BTE hearing aid and ear. ANSYS requires a contact between the hearing aid and ear. Defined contact surfaces are seen in Figure 7.4b. In Figure 7.5b the BTE hearing aid has a maximal displacement and in Figure 7.5c the BTE has been released. We assured that the hearing aid and ear were in contact after the displacement, and we introduced a stiffness of the ear of 4 MPa Youngs modul and 0.4 Poissons ratio, corresponding to a soft silicone material. Furthermore, we defined the contact as 'frictional' with a coefficient of 0.5, to keep the hearing aid from sliding off.

After the simulation, the deformed mesh of the ear and hearing aid was saved in Initial Graphics Exchange Specification (IGES) format and then re-imported



(a) Initial position of the BTE geometry. **(b)** Maximal BTE geometry displacement. **(c)** Final position of the BTE geometry.

Figure 7.5: A three step approach for BTE placement.

into the head, replacing the undeformed ear and hearing aid in it's original position. The updated geometry was then used for the acoustic simulation.

7.3.2 Harmonical simulations

A harmonical response analysis was used for simulating HRTFs for the two microphone positions microphone-ITE and BTE. Because the two cases have a similar approach, consisting of model setup, meshing and calculations, they are explained simultaneously. Differences between the cases will be emphasized.

7.3.2.1 Model setup

The two harmonical simulations were both based on a head geometry mounted on the torso geometry of the Brüel & Kjær HATS for the simulations. The microphone-ITE simulation used the CAD modified head model obtained in Chapter 6 and the BTE case was based on the head model with deflected pinna obtained in Section 7.3.1.

Our goal was to calculate the acoustical pressure in the air surrounding the head. The head model was placed inside a box with dimensions $420 \times 700 \times 250$ mm, which in the simulation was set to be filled with air. The model was subtracted from the box, leaving the air surrounding the model. The air-filled space is where the acoustical pressure occurs, and it is the interesting part of our model,

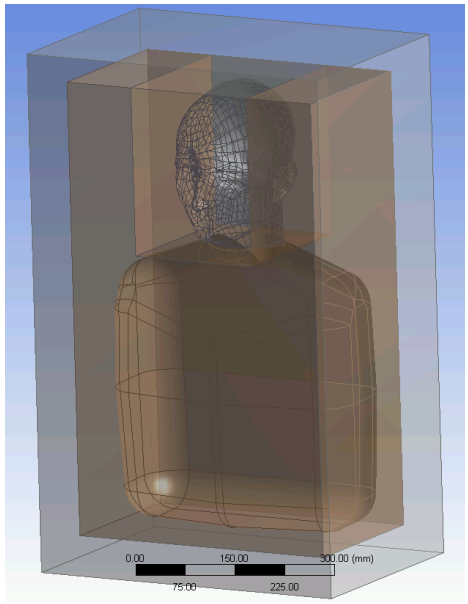


Figure 7.6: Geometry layers used for HRTF simulation. From inside to out: empty space shaped as the head and torso, inner air box, and perfectly matched layer.

equivalent to the physical world where sound reaches the ear through pressure changes in the air.

The acoustical measurements explained in Chapter 8 were executed in a semi-anechoic room in order to avoid room reflections. In the simulation a similar situation is created by adding a 40 mm perfectly matched layer (PML) to the outside of the inner air-box, see Figure 7.6. The purpose of the PML is to absorb radiated sound from the inner air part, and furthermore to enable calculations of sound pressure in the far field (outside the box). The default setting for attenuation in the PML region is used (0.001 or -60 dB), yielding 120 dB total attenuation of the sound field reflected from the PML (basically no reflections).

7.3.2.2 Meshing

It is not a trivial task to generate a mesh, which is fine enough without having so many nodes that calculation times becomes unmanageable. A rule of thumb when calculating acoustic pressure with first order elements is that the mesh should have at least 8 elements per wavelength. We would like to calculate

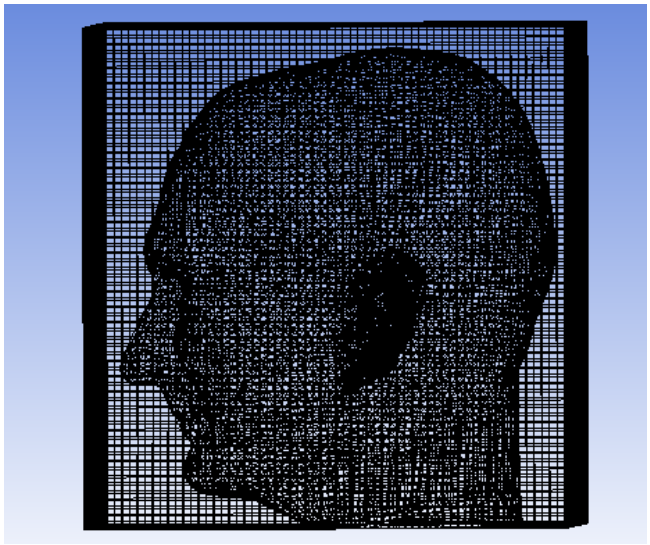


Figure 7.7: Mesh for the air surrounding the head.

HRTFs for up to 10,000 Hz, which then requires a mesh size smaller than: $\frac{343\ m/s}{10000\ Hz} \cdot \frac{1}{8} = 0.0043\ m$. The meshing approach differed for the microphone-ITE simulation and BTE simulation cases.

Meshing for the microphone-ITE simulation case: The head geometry was meshed with first-order acoustic elements (ANSYS type 30). The mesh was refined at the ears to an average edge length of 2 mm, while the rest of the model used a length of 4 mm. The re-meshing method was 'Hex dominant', which means that hexahedral elements were preferred. In regions where hexahedral elements were not flexible enough tetrahedral elements were employed instead. A mesh produced primarily from hexahedral elements contains less nodes compared to a pure tetrahedral mesh. The final mesh, which represented the model of the head and torso, contained 1.9 million nodes. Figure 7.7 shows the mesh for the air surrounding the head. Because we are dealing with a pure acoustical simulation one node contributes with one equation. If we had included vibrations on the surface of our model one node would contribute with four equations, one for pressure and three for displacement.

Meshing for the BTE simulation case: The head geometry was meshed with first-order acoustic elements (ANSYS type 30). Two meshes were created in order to reduce calculations times; one for frequencies $<$ 7.5 kHz and one for

frequencies < 10 kHz. The 7.5 kHz mesh only requires a mesh size smaller than $\frac{343\ m/s}{7500\ Hz} \cdot \frac{1}{8} = 0.0057\ m$.

Both meshes were refined at the ears to an average edge length of 1 mm. The remaining model had an edge length of 6 mm for the 7.5 kHz mesh and a length of 4 mm for the 10 kHz mesh. The re-meshing method was again 'Hex dominant' for both meshes. The 7.5 kHz mesh contained ~ 1.1 million nodes, whereas the 10 kHz mesh contained ~ 2.3 million nodes.

7.3.2.3 Calculations

In the acoustic measurement setup, see Chapter 8, loudspeakers surrounding the head generate the sound, and the microphones placed in the ear canal capture the sound pressure. In the simulation, the reciprocal approach is used [Mor86, ZDGG06]: the sound source is placed at the microphone position, and the sound pressure level is calculated on a sphere surrounding the head at 1.2 m distance. The reciprocal approach yields identical results to the direct approach, but the simulation can be performed for all sound-source directions in a single simulation step [JJ13].

7.4 Evaluation of simulations

Several considerations must be made concerning the accuracy and precision of the BTE placement simulation and the harmonical simulation.

The BTE placement was a very challenging task and it was therefore only performed once and only on the left ear. This opens up the question whether the simulated position of the BTE hearing aid is the correct or most optimal position. The simulated BTE position is to the best of our knowledge a realistic hearing aid placement. However, a hearing-aid user would reposition his or her hearing aid on a daily basis and the position is therefore varying. It would be interesting to simulate different hearing aid positions and investigate the effect on the simulated HRTFs, however due to time limitations this remains to be done. Kuhn and Burnett [KB77] showed that changing the position of a microphone placed behind the ear does have an effect on the sound pressure level, particularly at higher frequencies.

The major concern with the harmonical simulations is to keep the number of nodes low, in order to reduce calculation time, without loss of accuracy. The

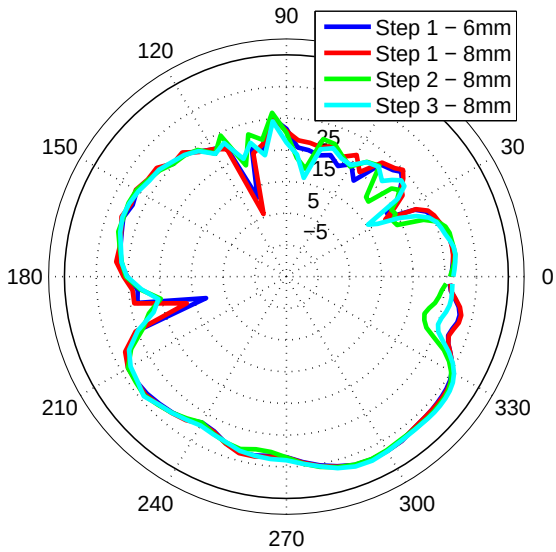


Figure 7.8: Comparison of HRTFs simulated for a frequency of 10 kHz for three torso cases and for two different element sizes. Step 1: no torso, Step 2: half of the torso and step 3: full torso. Element edge-sizes of 6 and 8 mm is tested.

number of nodes is a result of the element size for the mesh and the size of the model (head and torso). The model must at least contain the individual head model, however the importance of the torso is less known and it might be omitted or reduced to half size. Figure 7.8 investigates both mesh size and torso importance for a frequency of 10 kHz. 10 kHz is the highest frequency in the simulations and therefore the frequency where the smallest mesh size is needed. However, because the largest effect of the torso is observed at frequencies around 2-4 kHz, it would have been interesting to see the effect of the different torso cases at lower frequencies.

The figure reveals small differences between the mesh with an element edge-size of 6 and 8 mm. To be on the safe side, we choose to use a 4 mm mesh for a frequency of 10 kHz. The comparison of mesh sizes should therefore have contained for instance 3 and 4 mm, to assure that the 4 mm mesh is accurate enough. However, the process of generating a mesh is difficult and time consuming, and furthermore a mesh size of 3 mm would result in a mesh with a huge number of nodes and the calculation times would be very high. We therefore assume that a 4 mm mesh is adequate.

Figure 7.8 also reveals differences between the simulations of the different torso cases, step 1: no torso, Step 2: half of the torso and step 3: full torso. Because of differences between the cases we conclude that a full torso is required for our harmonical simulations, even though it is expensive in calculation times.

The simulated HRTFs need to be compared against equivalent measured HRTFs. This is the goal of Paper B and Chapter 9. The result and conclusion from the paper is that a similar overall behavior between the simulated and measured HRTFs are observed, however the simulations and measurements are definitely not equivalent. The final goal of this thesis is to optimize a directional microphone in a hearing aid based on the simulated HRTFs. The performance of a directional microphone in a BTE hearing aid optimized from simulated HRTFs is evaluated in Paper D. The paper shows that the directional microphone optimized from simulated HRTFs performs almost as well as the optimal directional microphone. It can therefore be concluded that our simulations have the required precision for the purpose of this thesis.

Chapter 8

Measured head related transfer functions

The focus of this chapter is to review the traditional methods used for measuring head related transfer functions (HRTFs), and to account for the HRTF measurements performed for this thesis. Two different measurement setups were used: A setup, that measures source positions on a sphere, referred to as ARI (Section 8.2.1) and a setup, that measures with source positions in the horizontal plane, referred to as Kongebakken (Section 8.2.2). The ARI measurements are the main data set used throughout this thesis. They consist of measurements for four human test subjects and a printed head. The Kongebakken measurement is a supplement consisting of measurements performed for the printed head and for the HATS. Section 8.3 provides an overview of the measured HRTFs.

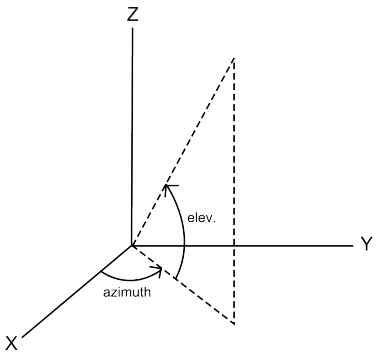
The ARI measurements were performed during a three-day measurement session with the participation of Piotr Majdak, Michael Mihocic, Martin Larsen, Rasmus R. Paulsen, Søren Laugesen, and the author. The primary measurement setup was accounted for by Piotr and Michael, whereas hearing-aid shells with microphones, connectors, and an amplifier was provided by Søren Laugesen and the author. The Kongebakken measurements were performed by Johannes Steininger at Oticon A/S in Smørum, Denmark.

The ARI measurements form the basis of Paper B, C and D, whereas the Kongebakken measurements are only used in Paper D.

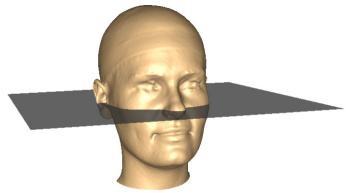
8.1 HRTF measurements

As stated in Chapter 2, head related transfer functions (HRTFs) are numerical representations of spatial acoustic cues. A HRTF is defined as the transfer function from a point in space to a microphone positioned in the ear canal normalized with a free-field response. HRTFs are individual, since they depend on head, pinna, and torso shapes.

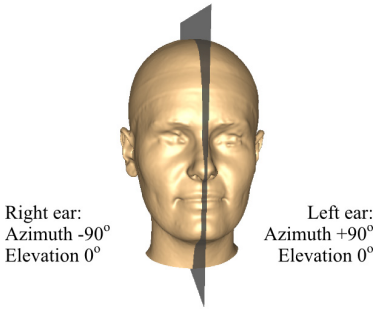
HRTFs are traditionally measured at a fixed radius from the center of the head to a source point, which are positioned with varying azimuth and elevation angles. Figure 8.1 defines the spatial coordinate system used in this thesis. The coordinate system is similar to what is used in much HRTF literature [CW99].



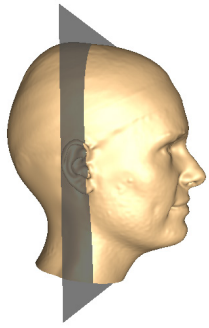
(a) Spherical coordinate system.



(b) Horizontal plane (elevation = 0°).



(c) Median plane (azimuth = 0°).



(d) Vertical plane (azimuth = 90°).

Figure 8.1: Definition of coordinate system.

HRTFs are generally measured for the left and right ear simultaneously. Micro-

phones are inserted partially into the ear canals. A known-spectrum stimulus is played at a given source position, defined by a radius, r, elevation θ, and azimuth ϕ. The stimulus could be a simple click, a pseudo-random binary sequences, or complementary Golay codes [CW99].

Usually, a set of HRTFs are measured. The set could consist of measurements in the horizontal plane at a number of azimuth angels. Such a set was measured in [NITI07] where HRTFs were measured for 72 azimuth angles spaced 5° apart, or in [WGS11] where the spacing was 1°. The set could also consist of 3D measurements with varying azimuth and varying elevation. In [DB07] a total of 449 HRTFs were measured with $\phi \in [0°;350°]$ and $\theta \in [-80°;90°]$, and in [GM+94] a total of 710 HRTFs were measured with $\phi \in [0°;355°]$ and $\theta \in [-40°;90°]$.

Since the spatial cues captured by HRTFs are essential for spatial hearing, it has been of great interest to investigate the structure of HRTFs. Cheng and Wakefield [CW99] show HRTF representations in the time, frequency, and spatial domains. Furthermore, they compare a set of measured HRTFs with HRTFs derived from a spherical head model. In this thesis, we will stick to the frequency domain, which is a popular domain for evaluation of peaks and notches in HRTFs [Bla83, KW92]. Furthermore, we also employ a spherical head model, as seen in Chapter 9.1.2, for comparison.

8.2 Our HRTF measurements

8.2.1 ARI measurements

The ARI HRTFs were measured using the measurement setup described in Paper B and the method described by Majdak et al. [MBL07]. The measurement setup is capable of measuring HRTFs at 3D source positions on a sphere, similar to measurement setups described in the literature [ADTA01, MSHJ95, DB07, GM+94].

The HRTFs were measured in a semi-anechoic chamber. The measurement setup is seen in Figure 8.2. Twenty-two loudspeakers were mounted on an arc at fixed elevations from −30° to 80°. Details are found in Paper B.

Measurements were performed both on real human subjects and on a printed 3D head model. Furthermore, a number of different microphone positions were measured, including positions on hearing aid shells. Section 8.3 provides an overview of the measurements.

Figure 8.2: Measurement setup with 22 loudspeakers and a test subject positioned in the center.

The human test subject or printed head was seated on a chair in the center of the arc and was wearing microphones or hearing aid shells. In either case the ear canals were blocked. Measurement frequencies ranged from 50 Hz to 20 kHz. The HRTFs were measured for one azimuth and several elevations at once by playing sweeps and recording the signals at the microphones. Then the subject was rotated by 2.5° to measure HRTFs for the next azimuth.

In the horizontal interaural plane, the HRTFs were measured with 2.5° spacing within the azimuth range of ±45° and with 5° spacing outside this range. In total, 1550 different source positions were measured, see Figure 8.3. The red points mark the source positions measured for the initial frontal position. The measurement procedure is constructed such that part of the source positions for the azimuth angle $\phi = 0°$ and part of the source positions for $\phi = 180°$, are recorded at the frontal position, while the other half is recorded after a 180° turn of the rotating chair. This procedure allows for a higher angular resolution for the elevation, since the loudspeaker distance can be twice as small, compared to if all loudspeaker positions were recorded at once. Notice in Figure 8.2, that

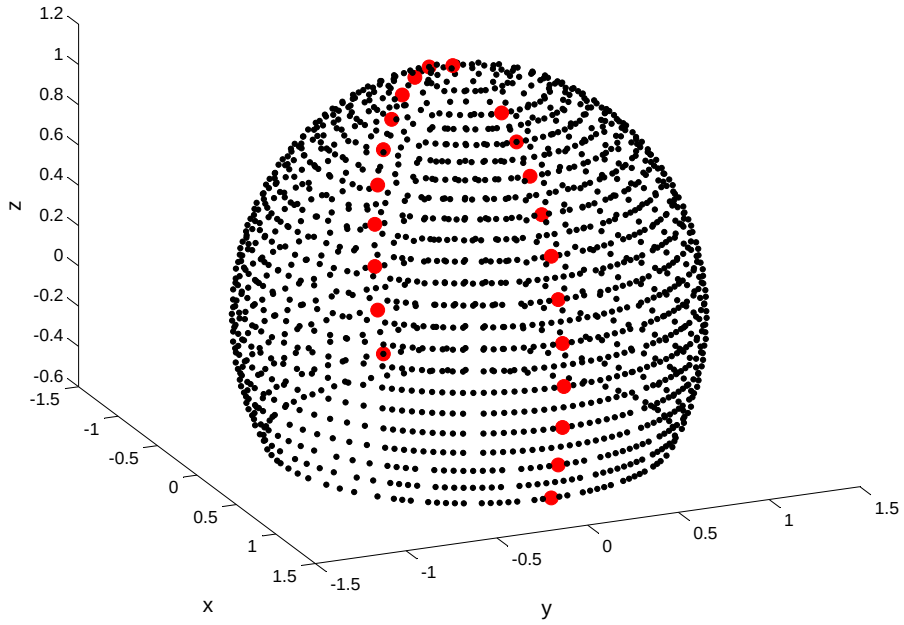


Figure 8.3: Source positions. Red points mark the source positions measured at the initial frontal position.

To decrease the total time required to measure the HRTFs, the multiple exponential sweep method (MESM) was applied [MBL07]. This method allows for a subsequent sweep to be played before the end of a previous sweep, but still reconstructs HRTFs without artifacts.

During the HRTF measurements, the position and orientation of the test subject's head were captured via an electromagnetic tracker (Flock of Birds, Ascension) in real time. The tracking sensor was mounted on the top of the test subject's (either human or printed head) head, see Figure 8.4. The left picture shows how the tracker was mounted, using double adhesive tape, on the printed head and the right picture show the mounting strips used for a human subject (no tracker present). The tracking device was capable of measuring 6 degrees of freedom (x, y, z, azimuth, elevation, and roll) at a rate of 51.5 measurements/sec. The tracking accuracy was 1.7 mm for positions and 0.5° for orientation. If the head was outside the valid range, the measurements for that particular azimuth were repeated once the subject was back in the range for 500 ms. The valid ranges were set to ± 2.5 cm for the position, ±2.5° for the az-

imuth, and $\pm 5^\circ$ for the elevation and roll. On average, measurements for three azimuths were repeated per subject and the complete measurement procedure lasted for approximately 20 minutes.

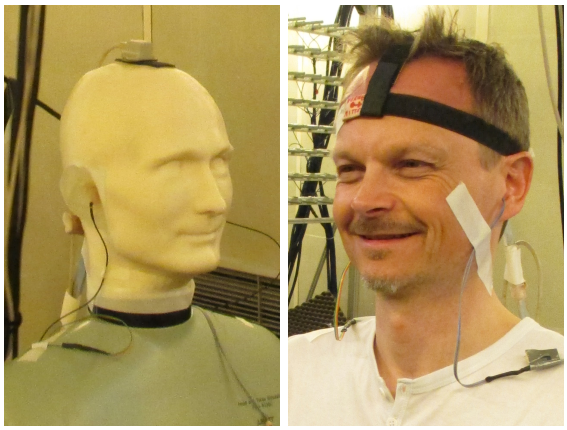


Figure 8.4: Tracker placement. Left: printed head. Right: human subject.

8.2.2 Kongebakken measurements

The Kongebakken HRTFs were measured in an anechoic room, using source positions in the horizontal plane. A single loudspeaker was positioned at a zero degree elevation. The printed head or HATS was placed on a turning table, and the table was rotated with 5° spacing for the azimuth angle.

A reference microphone (Brüel & Kjær 4191) was initially calibrated using a calibrator (1 kHz, 94 dB). The reference response was then measured with the microphone placed at a position corresponding to the position of the head, see Figure 8.5.

One ear was measured at a time with both CIC and BTE hearing aids. The opposite ear was occluded using tackywax. Measurement frequencies ranged from 94 Hz to 10 kHz. The HRTFs were measured for one azimuth at a time by playing a sweep and recording the signals at the microphones.

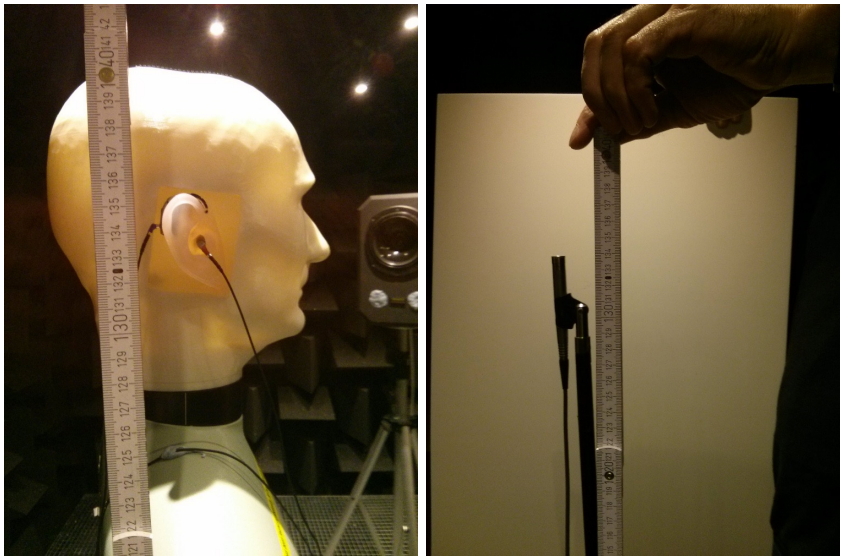


Figure 8.5: Left: measurement setup. Right: placement of reference microphone.

8.3 Overview of HRTF measurements

A number of different HRTFs were measured for this thesis. The HRTFs vary both with respect to microphone position, test subject (human or head and torso simulator) and measurement setup (ARI or Kongebakken).

Four different microphone configurations were used:

- A single microphone placed in the ear canal, Figure 8.6a.
- Two microphones incorporated into a BTE shell, Figure 8.6b.
- A single microphone incorporated into a CIC shell, Figure 8.6b.
- Two microphones incorporated into an ITE shell, Figure 8.6c.

Four human test subjects (NH 166, NH 167, NH 168 and NH 170) and a printed head model (NH 169) of subject NH 167, see Chapter 6, were measured using the ARI setup. Furthermore, a remeasurement of subject NH167 was obtained. This measurement is referred to as NH167-2. We follow the naming convention from the ARI database [ARI], where data is available. The printed head and

the HATS were measured using the Kongebakken measurement setup. Table 8.1 provides an overview of the measurements.

(a) Mic. in ear canal. **(b)** BTE and CIC shells. **(c)** ITE shell.

Figure 8.6: Microphone positions.

	Mic.		BTE		ITE		CIC	
	L	R	L (F/R)	R (F/R)	L (F/R)	R (F/R)	L	R
ARI								
NH166	×1	×1	×1	×1	×1	×1	×1	×1
NH167	×2	×2	×2	×2	×2	×2	×1	×1
NH168	×1	×1	×1	×1	×1	×1	×1	×1
NH169	×1	×1	×1	×1	×1	×1	×1	×1
NH170	×1	×1	×1	×1	-	-	-	-
Kongebakken								
HATS	-	-	×1	×1	-	-	×1	×1
NH169	-	-	×1	×5	-	-	×1	×5

Table 8.1: Overview of HRTF measurements. The ARI data is accessible from [ARI]. NH169 is the printed head mounted on the HATS torso.

8.3.1 A reliability study

The variation between measured HRTFs is highly relevant, since this variation forms the basis for comparing HRTFs. A set of repeated measurements were therefore measured for the BTE and CIC shells in the ARI measurement setup. Paper C presents the reliability study in details. The measurements

were performed using the printed head, test subject NH169, mounted on a torso simulator. Three different conditions were recorded:

- **Control:** The measurements were repeated without any intervention.
- **HAD:** Hearing aids were detached from the head and re-attached before the next measurement.
- **Tripod:** The tripod carrying the torso, head, and the hearing aids was removed from the arc center and placed back. Furthermore, hearing aids were detached from the head and re-attached.

As seen in Table 8.2 all measurements were repeated 5 times. The HAD condition is missing for the CIC, because of difficulties with repositioning the hearing aid in the ear canal. The printed ear canal lacked the flexibility that a human ear canal has, which made it difficult to insert the hearing aid. The CIC Tripod case therefore only contains repositioning of the torso and head, and not of the CIC.

	BTE		CIC	
	L (F/R)	R (F/R)	L	R
ARI				
NH169 - Control	×5	×5	×5	×5
NH169 - HAD	×5	×5	-	-
NH169 - Tripod	×5	×5	×5	×5

Table 8.2: Overview of HRTF measurements for the reliability study.

The measured data sets are evaluated in the next chapter.

Chapter 9

Evaluation of HRTFs

This chapter contains a visualization and evaluation of synthesized, measured and simulated HRTFs. Furthermore, post-processing procedures for the measured and simulated data are accounted for. The work described in this chapter consists of synthesizing free-field responses and spherical head-model HRTFs and of performing post-processing of the ARI measurements. Besides that, this chapter mainly contains a visualization and discussion of the simulated and measured HRTFs, which were explained in Chapter 7 and 8, respectively.

The free-field response and spherical head-model HRTFs were synthesized by the author, using the implementation of a spherical head-model developed by Søren Laugensen et al. [LRC03]. The ARI measurements were post-processed using a procedure developed in collaboration with Søren Laugesen, and implemented by the author.

Measured and simulated HRTFs are also presented in Paper B and C, and utilized in Paper D. The present chapter presents HRTFs measured for BTE and ITE hearing aids using the ARI setup, along with HRTFs simulated for a BTE hearing aid. Paper B, furthermore, presents simulated and measured HRTFs for a microphone placed in the ear canal.

9.1 Synthesized HRTFs

We synthesize both free-field responses and HRTFs for a spherical head-model. The free-field responses are important because their behavior is known and can be described analytically. As an example we use the free-field responses to test the weighting scheme developed in Paper D, because DI values exist as a ground truth. The behavior of the HRTFs become more complicated when a sphere is introduced into the sound field and analytically derived ground truth values does not exist. However, a sphere is still a more simple geometry than a human head, which has both pinna and a torso to complicate the sound field. The spherical head-model is therefore used as a simple case that can be compared to the more complicated human HRTFs.

9.1.1 Free-field model

Free-field responses are synthesized for two microphones, simulating a BTE or an ITE hearing aid. The front microphone is placed in the center of the coordinate system and the rear microphone is placed 10 mm away along the negative x-direction. Free-field responses, C_{FF}, are calculated for each frequency, f, and source position, (θ, ϕ), individually. The front microphone response is assumed to be:

$$C_{FF_f}(f, \theta, \phi) = 1 \tag{9.1}$$

and the rear microphone response then becomes:

$$C_{FF_r}(f, \theta, \phi) = \exp\left(\frac{-i2\pi f d \cos(\rho(\theta, \phi))}{c}\right), \tag{9.2}$$

where d is the distance from the center of the coordinate system along the x-axis, $\rho(\theta, \phi)$ is the angle between the x-axis and a given source position (θ, ϕ).

9.1.2 Spherical head-model

HRTFs are synthesized for a set of microphones placed on the side of a sphere with a diameter of 17.5 cm. The front microphone is placed on the y-axis lifted 5 mm from the surface of the sphere. The rear microphone is placed 10 mm away in the negative x-direction. The spherical head-model is deduced in Duda and Martens [DM98] and we use the implementation developed by Laugesen et al. [LRC03].

Figure 9.1 shows magnitude and phase for the free-field responses and spherical head-model HRTFs for the frontal position (elevation, azimuth) = (0,0), which means that the sound is arriving from the front. The magnitude is constant in the free-field case, whereas the effect of the sphere is seen for frequencies above 1 kHz. Frequencies below 1 kHz are unaffected by the presence of the sphere because the wavelength of the sound is much larger than the diameter of the sphere. The phase plot shows a phase delay between the front and rear microphones which is similar for the free-field and spherical head-model case. An equivalent phase delay was expected because the distance between the microphones is identical for the two data sets.

Figure 9.2 shows polar plots of the magnitude for frequencies of 1, 2, 4 and 8 kHz for the spherical head-model case. The polar plots reveal a shadowing effect of the contralateral side. The shadowing effect is present for frequencies with a wavelength that is comparable to or smaller than the size of the sphere. Notice also the peak in magnitude at approximately 270°, which is the point where sound traveling opposite ways around the sphere meet.

9.2 The ARI reliability study

The ARI reliability measurements were explained in Section 8.3.1 and is used in Paper C to evaluate the reliability of measured HRTFs. The HRTFs were measured using a BTE hearing aid, which is one of the hearing aid types that will be used in Chapter 10 to optimize directivity. The reliability study is therefore highly relevant.

The measured HRTFs were post-processed to remove room and loudspeaker responses. The HRTFs also contain responses from the particular microphone that was used, however our attempt to remove individual microphone responses removed the phase difference between the set of BTE microphones. We therefore assume similar microphone responses between a matched pair of BTE microphones, and utilize the front microphone to compensate for both microphones.

For a given microphone k, a compensated HRTF $C_k(\theta, \phi)$ at elevation θ, and azimuth ϕ, is calculated in the frequency domain:

$$C_k(\theta, \phi) = \frac{H_k(\theta, \phi)}{E_F(\theta)}, \tag{9.3}$$

where $H_k(\theta)$ is the windowed raw HRTF for microphone k and $E_F(\theta)$ is the

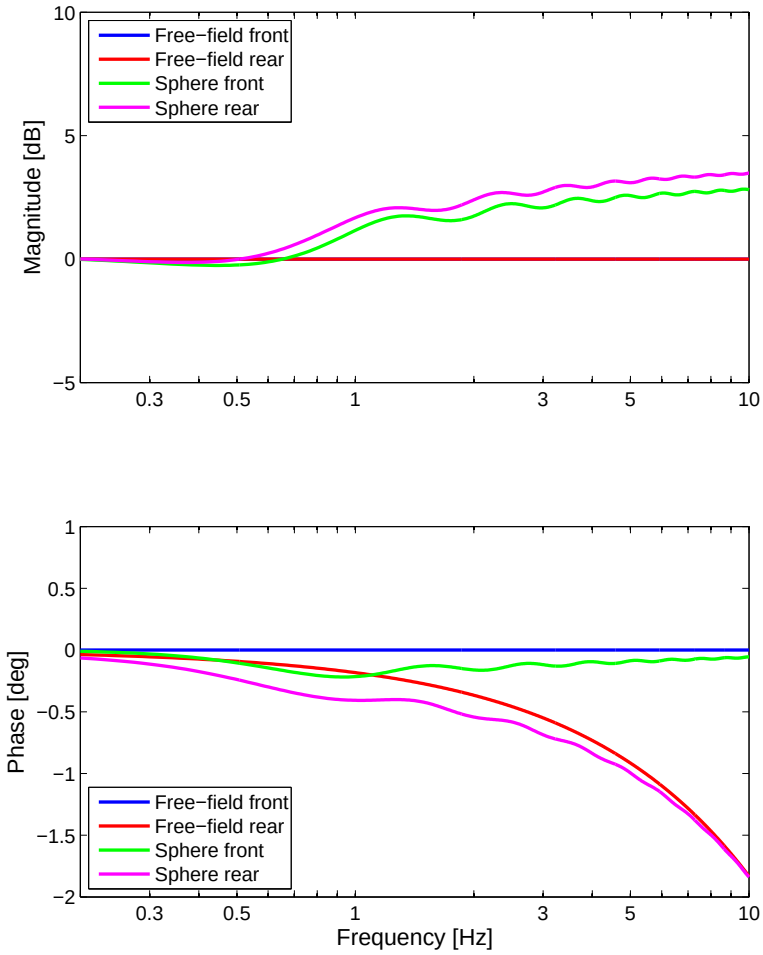


Figure 9.1: Synthesized free-field responses and spherical head-model HRTFs: Magnitude (top) and phase (bottom) for the front and rear microphones at the source position (elevation, azimuth) = (0,0).

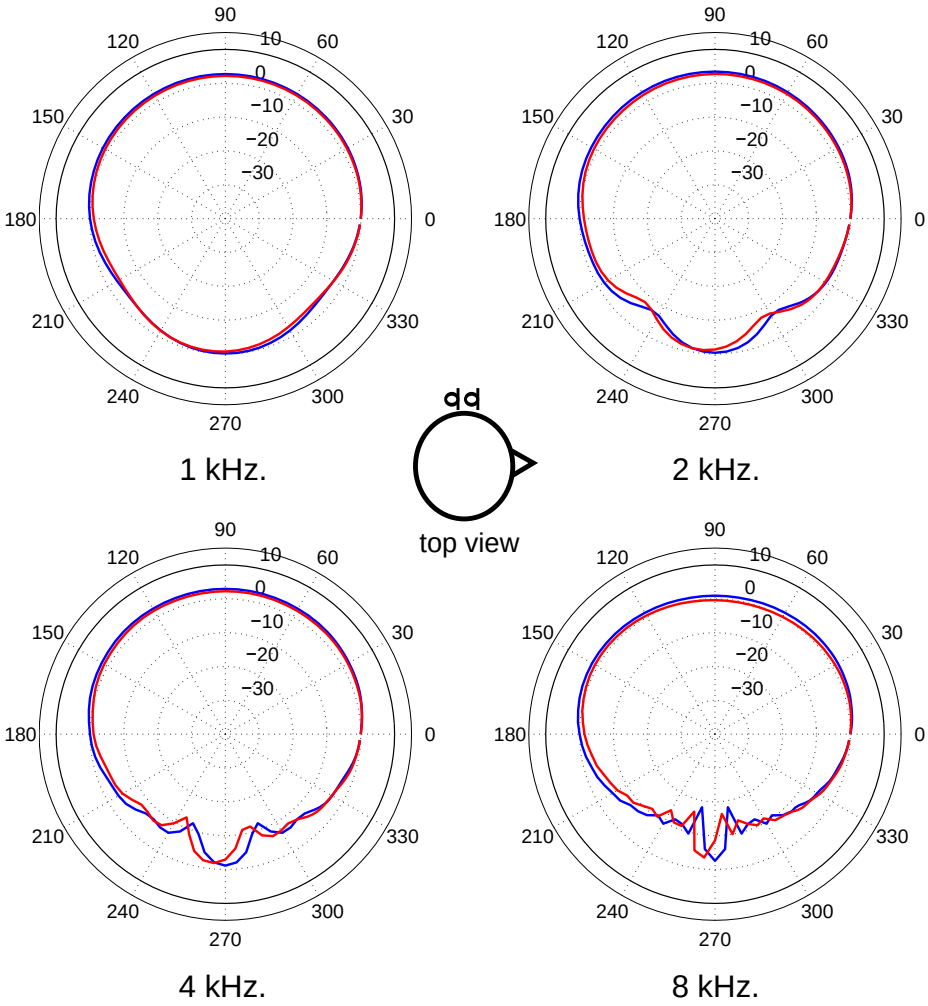


Figure 9.2: Synthesized spherical head-model HRTFs: Magnitude for HRTFs in the horizontal plane (elevation = 0). The front and rear microphones are placed on the left side of the spherical head-model. Magnitude is given in dB along the radial axis.

windowed equipment transfer function (ETF) for the front microphone in the BTE hearing aid.

Three different cases were measured: Control, HAD and Tripod. Recall that no changes were made to the setup for the Control case, that the BTE was repositioned between measurements for the HAD case and that the Tripod case both contained a repositioning of torso and head, and a repositioning of the BTE.

Figure 9.3 shows five repeated measurement for the Control case. The top of the figure contains the magnitude, while the bottom figure contains the phase. Only the magnitude was evaluated in Paper C. The phase is however very important, because the HRTF measurements will be used for directivity calculations, which are highly dependent on phase differences between the front and rear microphone. Magnitude and phase appear very accurate among measurements for both the front and rear microphone.

Figure 9.4 shows five repeated measurement for the HAD case. The five measurements show a higher variation for both magnitude and phase. It is however reassuring to see that the phase difference within a microphone pair seems consistent between measurements.

Figure 9.5 shows repeated measurements for the Tripod case. The measurements appear less accurate, however again the phase difference between a microphone pair is maintained.

Notice how well the phase differences between the front and rear microphones are maintained, even though the HRTFs have been compensated for room reflections, loudspeaker responses, and for a general microphone response. This is due to the post-processing procedure where both front and rear microphones are compensated using a reference measure for the front microphone.

9.3 Kongebakken - repeated measurements

The repeated HRTF measurements performed using the Kongebakken setup are less interesting than the ARI reliability study, since the measurements were performed using a CIC placed in the ear canal instead of a BTE or ITE hearing aid. The repeated measurements are however valuable for assuring that the measurements are consistent. The CIC hearing aid was repositioned between each repeated measurement. The measured HRTFs were neither compensated for loudspeaker (since only a single loudspeaker was used) nor microphone re-

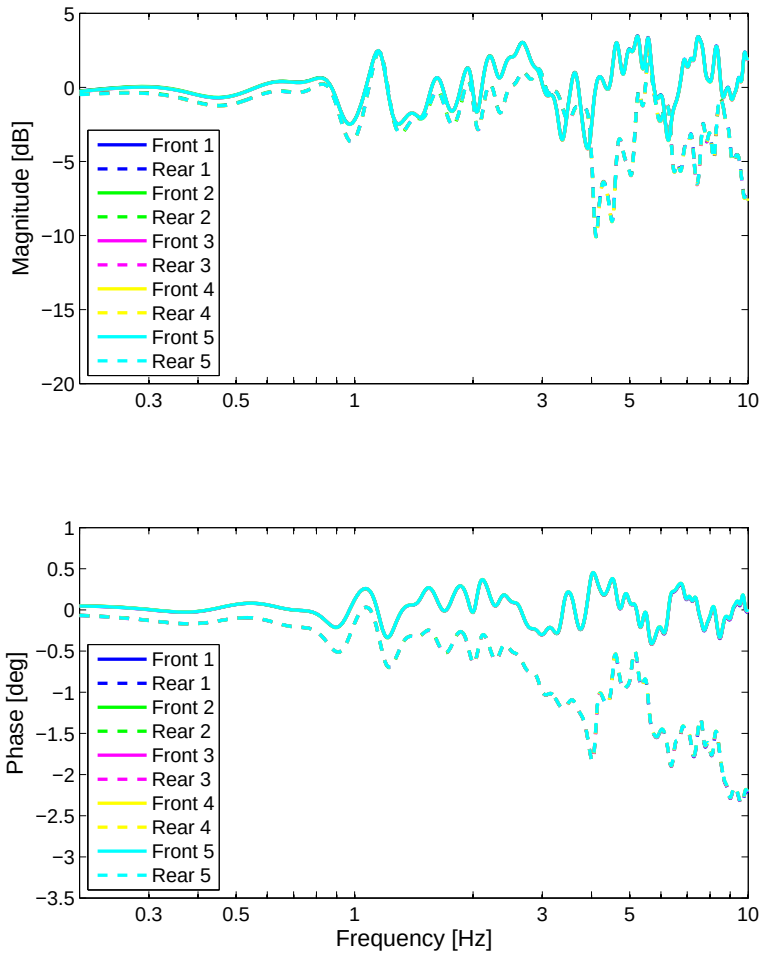


Figure 9.3: ARI measured HRTFs for a BTE hearing aid: Magnitude (top) and phase (bottom) for 5 repeated measurements of the Control case. Front and rear microphones are shown for a source position of (elevation, azimuth) = (0,0).

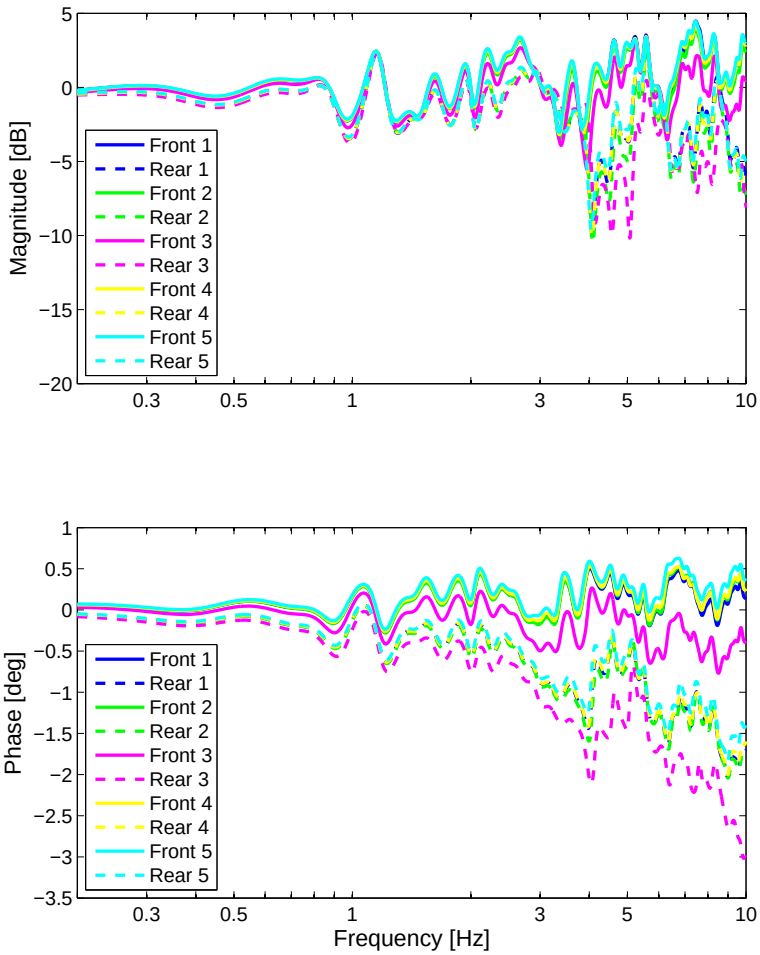


Figure 9.4: ARI measured HRTFs for a BTE hearing aid: Magnitude (top) and phase (bottom) for 5 repeated measurements of the HAD case. Front and rear microphones are shown for a source position of (elevation, azimuth) = (0,0).

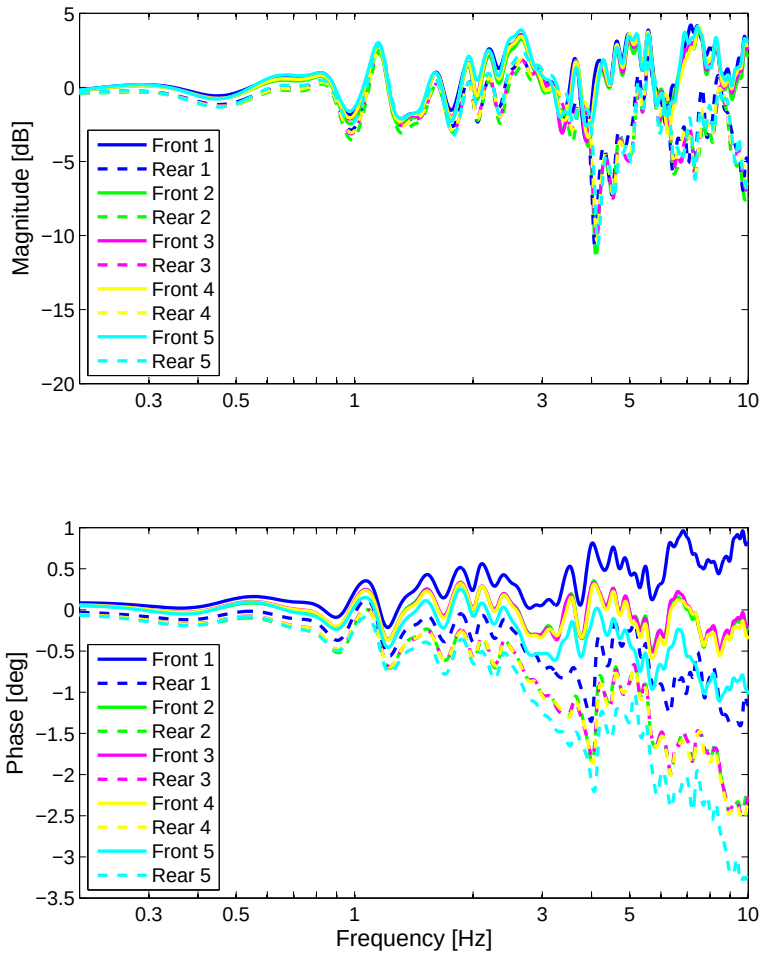


Figure 9.5: ARI measured HRTFs for a BTE hearing aid: Magnitude (top) and phase (bottom) for 5 repeated measurements of the Tripod case. Front and rear microphones are shown for a source position of (elevation, azimuth) = (0,0).

sponses, however they were time-shifted 3.9 ms in order to remove the internal system delay and achieve a relatively constant phase. Figure 9.6 shows that both magnitude and phase are measured relatively reliable, and in a range comparable to what was shown for the ARI HAD case in Figure 9.4.

The present set of repeated measurements were, unlike the ARI reliability data, measured for all azimuth angles. It is therefore interesting to evaluate how accurate the measurements are for a given frequency as a function of azimuth position. Figure 9.7 shows polar plots of the magnitudes for frequencies of 1, 2, 4 and 8 kHz. The figure reveals relatively accurate magnitudes for all azimuth angles, however the accuracy seems to drop for higher frequencies. This is probably due to limitations given by the angular step size.

9.4 Comparison between different HRTFs

Paper B provides a comparison between measured and simulated HRTFs for a microphone placed in the ear canal. The paper reveals that the simulated HRTFs correlate, to some extend, with the measured HRTFs. The paper is only concerned with comparing magnitudes.

The HRTFs that will form the basis for optimization of the directional microphone in Chapter 10 are HRTFs measured and simulated for microphone pairs in a BTE hearing aid and HRTFs measured for an ITE hearing aid. Figure 9.8 shows magnitude and phase for microphone pairs in a BTE hearing aid, for the simulated and the two measured sets of HRTFs. The ARI HRTF set was normalized as in Equation 9.3. The frontal position (elevation, azimuth) = (0,0) is used with offset magnitudes, to achieve HRTFs roughly in the same magnitude range. Kongebakken HRTFs are, furthermore, time-shifted 3.9 ms to remove the internal system delay. The magnitude and phase differences observed for the Kongebakken data arise because the dataset are not normalized with a free field response as the simulated and ARI HRTFs. However, both the ARI measured HRTFs and the simulated HRTFs are normalized with a free field response, why both magnitudes and phases ought to be in the same range. The simulated HRTFs do however contain a magnitude offset which is not fully understood.

It is obvious that the three data sets are not equivalent regarding either magnitude or phase. The magnitudes do, however, have a similar general behavior. The ARI HRTFs have more fluctuations compared to the other two data sets, which is probably the result of a finer frequency resolution than for the two other cases. Furthermore, it is noticed that the HRTFs measured with the Kongebakken setup have a roll-off at lower frequencies. The roll-off originate

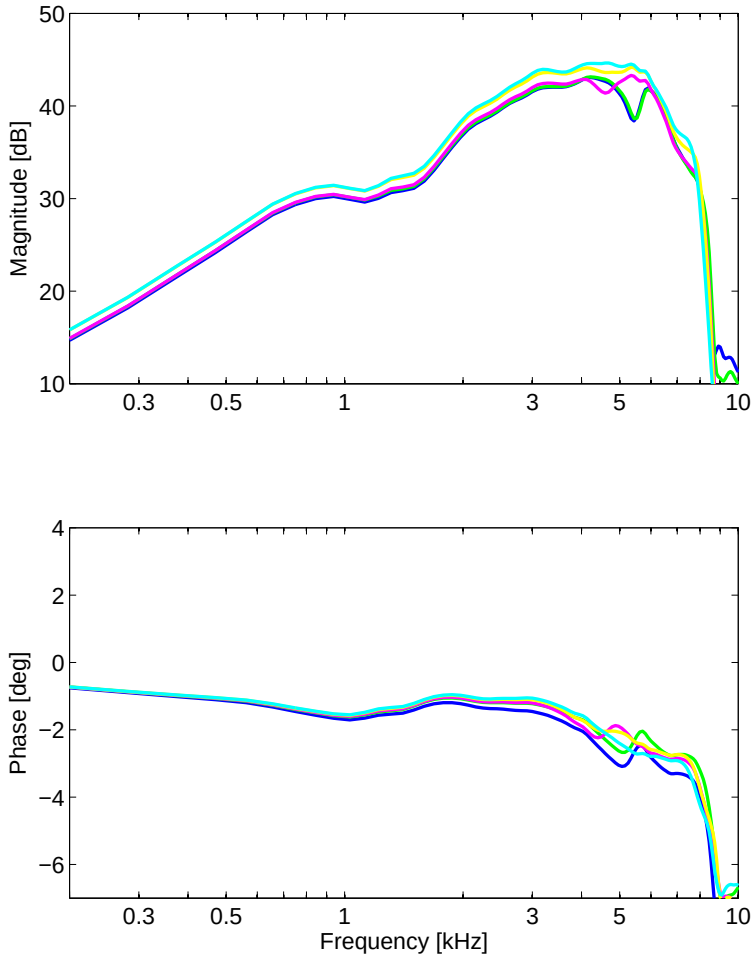


Figure 9.6: Kongebakken measured HRTFs for a CIC hearing aid: Magnitude (top) and phase (bottom) for 5 repeated measurements. Hearing aids were repositioned between each repeated measurement. A source position of (elevation, azimuth) = (0,0) was used. The HRTFs were time-shifted 3.9 ms to compensate for the internal system delay.

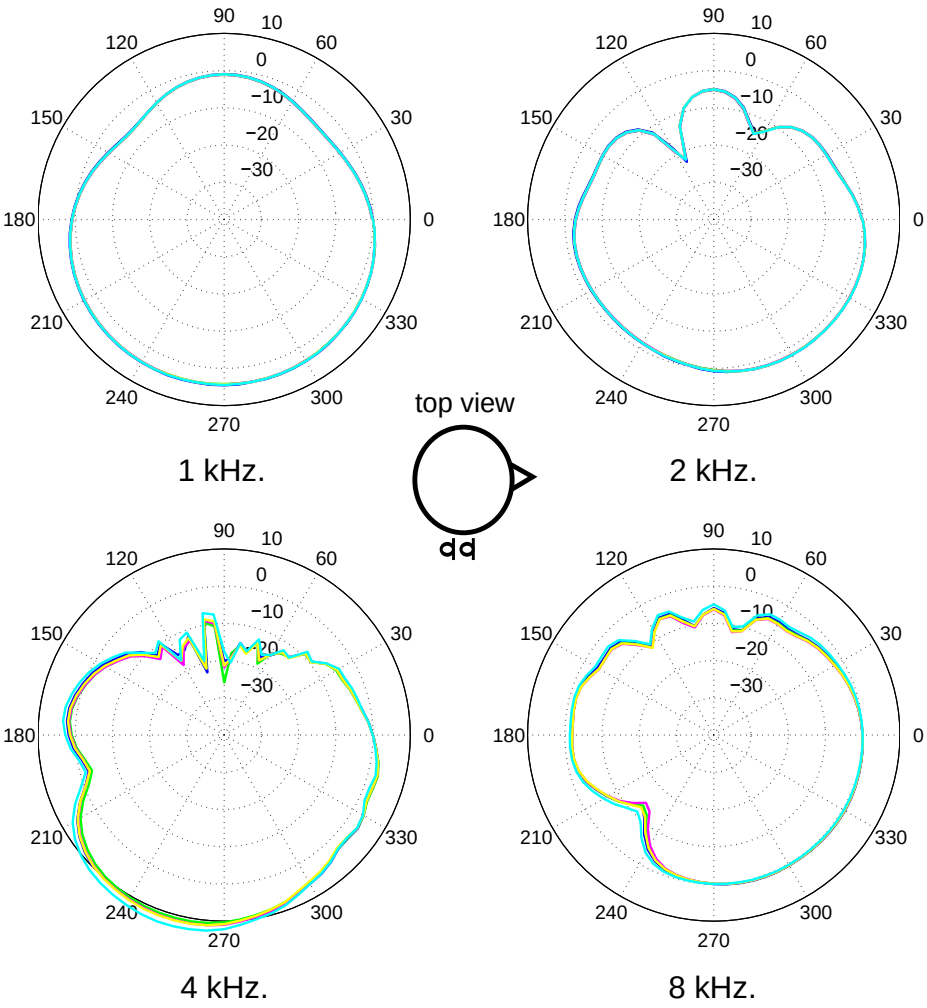


Figure 9.7: Kongebakken measured HRTFs for a CIC hearing aid: Magnitude for HRTFs in the horizontal plane (elevation = 0) for 5 repeated measurements. Hearing aids were repositioned between each repeated measurement. The microphone is placed on the right side of the head. Magnitude is given in dB along the radial axis.

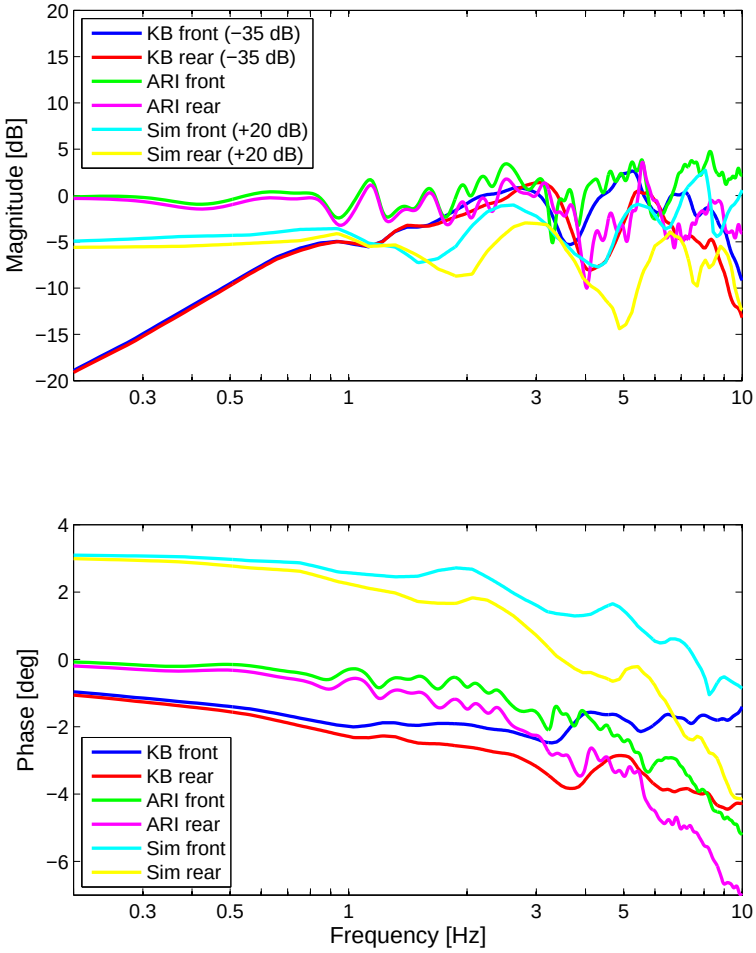


Figure 9.8: Magnitude (top) and phase (bottom) for ARI and Kongebakken (KB) measured HRTFs, and for simulated (Sim) HRTFs. All three cases were obtained using a BTE hearing aid. The source position is (elevation, azimuth) = (0,0). Kongebakken HRTFs are time-shifted 3.9 ms to remove the internal system delay.

from the response of the microphones in the BTE. The reason that the roll-off is not present in the ARI HRTFs is that the ARI data have been compensated for a general microphone response in the post-processing procedure in Equation 9.3.

9.5 Comparison between BTE and ITE hearing aids

HRTFs were measured for both BTE and ITE hearing aids using the ARI measurement setup and both kinds of hearing aids are used for optimization of the directional microphone in Chapter 10. It is therefore interesting to compare the two HRTF sets. The main difference between the two hearing-aid types is the position of the hearing aids and thereby the position of the microphones on the ear. The BTE hearing aid is placed on behind the ear and the ITE is placed in the ear canal. The position of the microphones on the head affects the microphone's directional response. Figure 9.9 shows the magnitudes for the front microphone in a BTE and an ITE hearing aid in the horizontal plane. The highest magnitudes are located approximately at a 40° azimuth for the ITE hearing aid for frequencies of 2 and 4 kHz, whereas the BTE hearing aid has its highest sensitivity for azimuth angles ranging from 30° to 180°. Differences in the directional response between the BTE and ITE hearing aids are most simple to observe for the two center frequencies. The shadowing effect of the head and particularly the effect of pinna is not observed for a frequency as low as 1 kHz, because the wavelengths are longer than the size of the head. Furthermore, the sound field becomes complicated for frequencies as high as 8 kHz, probably because of reflections and diffractions around the pinna and inside the concha. The position of the BTE hearing aid behind the ear results in a shadowing effect, caused by the pinna, for sounds arriving from the front. The position of the ITE hearing aid in the ear canal results in a shadowing effect, caused by pinna, for sounds arriving from the rear. The largest pinna effect is therefore obtained for an azimuth angle of 40°, and not for the front position. This observation will become important in Chapter 10. Similar results were obtained in [Dil12, RM99].

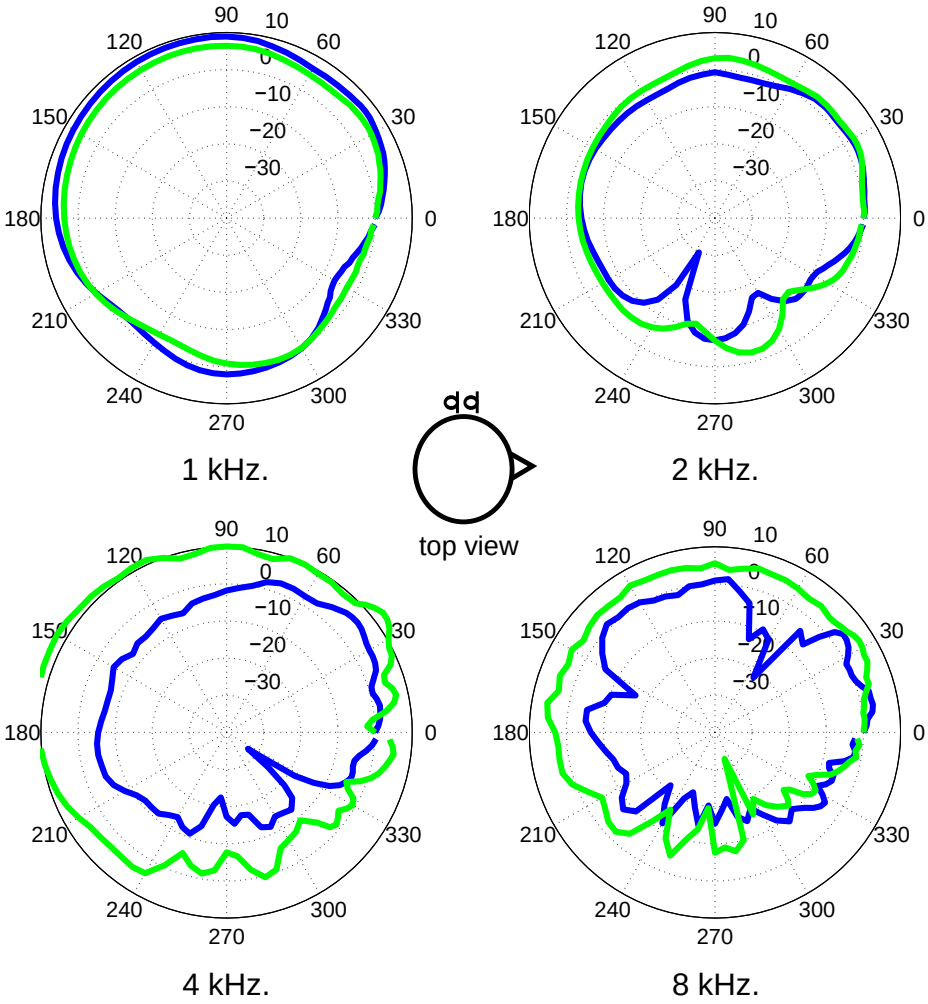


Figure 9.9: ARI measured HRTFs for the front microphone in a BTE (green line) and an ITE (blue line) hearing aid: Magnitude for HRTFs in the horizontal plane (elevation = 0) measured with the hearing aids placed on the left side of the head. Magnitude is given in dB along the radial axis.

Chapter 10

Optimization of the directional microphone

In this chapter we propose a method for listener-specific optimization of the directional filters in a BTE and an ITE hearing aid. The method consists of two parts: a weighting scheme to account for a non-uniform distribution of source positions and an optimization scheme. The chapter also contains the main result of this thesis, namely a verification of how well directional filters optimized on individually simulated HRTFs perform on the individual human subject. The later results are only obtained for a BTE hearing aid.

The proposed weighting scheme was developed by Søren Laugesen and the author. The optimization scheme was conceived by Michael S. Pedersen and the author. Both methods were implemented by the author.

The present chapter has a relatively high degree of overlap with Paper D, since the focus of the paper is similar to the goal of this thesis.

10.1 Directivity index

Recall from Chapter 3 that different measures exist for improving directionality in a hearing aid, such as the front-back ratio, the unidirectional index (UI), and the directivity index (DI). As stated in [DB07] the DI is a useful measure that explains the sensitivity for frontal sounds relative to an average across all directions. The DI is however very sensitive towards the front microphone measurement because a single measurement is used to estimate sounds from the front, instead of using an average over the front positions as it is the case for the UI. The DI has traditionally been a popular measure for directionality and was the preferred measure in papers like [DB07] and [SR93]. In keeping with tradition we employ the DI in this thesis.

The directivity index (DI) is in [DB07] defined as:

$$DI(f) = 10\log_{10}\left[\frac{4\pi|p(0,0)|^2}{\int_0^{2\pi}\int_0^{\pi}|p(\theta,\phi)|^2|sin(\theta)|d\theta d\phi}\right], \tag{10.1}$$

where $|p(0,0)|^2$ is the magnitude of the on-axis mean-square sound pressure microphone response to a plane wave in free field, and $|p(\theta,\phi)|^2$ is the magnitude of the spatially averaged mean-square sound pressure microphone response.

As explained in Dittberner [DB07] the on-axis free-field response of a microphone is easily computed. However, the spatially averaged response is not easily computed. Two methods exist for computing the spatially averaged response. The first method is referred to as the *diffuse field method*, where a microphone is placed in an artificially generated diffuse sound field and the response is measured. The second method is referred to as the *free-field method*, where the microphone is placed in free field and the response from a sound source is measured successively for many positions on a sphere. Using the free-field method the DI would be calculated as:

$$DI(f) = 10\log_{10}\left[\frac{4\pi\cdot|p(0,0)|^2}{\sum_{m=1}^{\frac{2\pi}{\Delta\phi}}\sum_{n=1}^{\frac{\pi}{\Delta\theta}}|p(\theta_n,\phi_m)|^2|sin(\theta_n)|\Delta\theta\Delta\phi}\right], \tag{10.2}$$

where $\Delta\theta$ and $\Delta\phi$ are stepsizes for elevation and azimuth angle, respectively. The spatially averaged response is estimated from a discrete number of source positions that should be evenly distributed on a sphere surrounding the head.

10.2 Calculating DI using non-uniformly sampled source positions

HRTFs are often acoustically measured using measurement setups as the ones explained in Chapter 8. With respect to the DI, having a limited number of source positions means that Equation 10.2 has to be used. Furthermore, the distribution of source positions might not be uniformly distributed on the sphere.

Our source positions are non-uniformly distributed on the sphere. We lack source positions below elevations of $-30\,^\circ$ and furthermore the frontal part is more densely sampled, see Figure 8.3.

We developed a weighting scheme, inspired by the method described by Dittberner et al. [DB07], to compensate for non-uniformly distributed source positions. The principle is to calculate a weight for each source point based on how large a fraction of the spheres surface area that the source point accounts for.

The surface area of a sphere slice covered by the ith elevation θ_i is:

$$S_E(\theta_i) = \int_{l_i}^{u_i} 2\pi g(z)\sqrt{1+[g'(z)]^2}dz, \tag{10.3}$$

where $g = \sqrt{1-z^2}$, $u_i = \theta_i + \left(\frac{\theta_{(i+1)}-\theta_i}{2}\right)$ and $l_i = \theta_i - \left(\frac{\theta_i-\theta_{(i-1)}}{2}\right)$. Here it is assumed that g revolves around the z-axis and that the radius of the sphere is one (the radius is insignificant, since only ratios are considered). The boundaries u_i and l_i are positions on the z-axis and are calculated as the sine of the angular-boundaries of a given elevation.

The surface area, S_{EA}, for the ith elevation and jth azimuth angle is now given by:

$$S_{EA}(\theta_i,\phi_j) = \frac{(\phi_{j+1}-\phi_{j-1})/2}{2\pi}\cdot S_E(\theta_i). \tag{10.4}$$

Each measurement point is weighed according to the size of the surface area that it covers, using the weights:

$$\alpha(\theta,\phi) = \frac{S_{EA}(\theta,\phi)}{4\pi}. \tag{10.5}$$

Using an α-weight for each source position Equation 10.2 becomes:

$$DI(f) = 10\log_{10}\left[\frac{4\pi|C(0,0)|^2}{\sum_{m=1}^{M}|C(\theta_m,\phi_m)|^2\cdot\alpha(\theta_m,\phi_m)}\right], \tag{10.6}$$

where (θ_m, ϕ_m) is the m'th source position and where M is the total number of source positions. The proposed DI weighting scheme will be evaluated using a set of synthesized free-field responses in Section 10.4.

10.3 DI optimization of a directional microphone

The principle of a directional microphone was explained in Chapter 3. The response from a directional microphone, for a single frequency, is:

$$C_D(\theta, \phi) = C_F(\theta, \phi)w_1 + C_R(\theta, \phi)w_2, \tag{10.7}$$

where $C_F(\theta, \phi)$ and $C_R(\theta, \phi)$ is the HRTFs for the front and rear microphone, respectively, and w_1 and w_2 are complex weights. Notice that the complex weights are calculated for each frequency independently.

Using matrix notation:

$$\mathbf{C}_D = \mathbf{C}\mathbf{w}, \tag{10.8}$$

where

$$\mathbf{C} = \begin{bmatrix} C_F(0,0) & C_R(0,0) \\ \vdots & \vdots \\ C_F(\theta_m, \phi_m) & C_R(\theta_m, \phi_m) \end{bmatrix}$$

and

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

If the weight $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ we simply obtain the front-omni response. The expressions for other standard filters, such as a cardioid and a hyper-cardioid, are stated in Chapter 3.

Equation 10.6 is optimized using a directional microphone as defined in Equation 10.7 and 10.8:

$$DI(f) = 10\log_{10}\left[\frac{4\pi|C_D(0,0)|^2}{\sum_{m=1}^{M}|C_D(\theta_m, \phi_m)|^2 \cdot \alpha(\theta_m, \phi_m)}\right] \propto \frac{(\mathbf{C}_0\mathbf{w})^H(\mathbf{C}_0\mathbf{w})}{(\mathbf{C}_\alpha\mathbf{w})^H(\mathbf{C}_\alpha\mathbf{w})}, \tag{10.9}$$

where ^{H} denotes the complex conjugate transpose,

$$\mathbf{C}_0 = [C_F(0,0)\ C_R(0,0)]$$

and,

$$\mathbf{C}_\alpha = \begin{bmatrix} C_F(0,0)\sqrt{\alpha(0,0)} & C_R(0,0)\sqrt{\alpha(0,0)} \\ \vdots & \vdots \\ C_F(\theta_m,\phi_m)\sqrt{\alpha(\theta_m,\phi_m)} & C_R(\theta_m,\phi_m)\sqrt{\alpha(\theta_m,\phi_m)} \end{bmatrix}.$$

The cost function in Equation 10.9 is similar to the cost functions proposed by [SR93, GB00] and it can be optimized using the generalized eigenvalue method.

10.4 Evaluation of the optimization scheme

The DI method described in Equation 10.6 is evaluated using a set of synthesized free-field responses, see Section 9.1.1. The HRTFs are synthesized as two microphones placed 10 mm apart, resembling a BTE or ITE hearing aid. Two sets of source positions are evaluated. The first set consists of the 1550 source positions used in our acoustical measurements in Chapter 8. The second set consists of the 1550 source positions plus an additional 448 source positions to cover the missing data on the lower part of the sphere, resulting in 1998 positions. The two set of HRTFs are from now on referred to as Syn_{1550} and Syn_{1998}.

Equation 10.6 provides DI values as a function of frequency. It is common practice to average over frequencies to obtain a single DI value [DB07]. However, it is not the entire frequency range that is important for speech intelligibility. The articulation index (AI) [DB07] is a frequency weighting scheme where frequencies that are important for speech are weighted high. We calculate an Articulation Index weighted Directivity Index (AI-DI) for front-omni, cardioid and optimal filters for both Syn_{1998} and Syn_{1550}. The values are seen in Table 10.1. The table also contains 'asymptotical' DI (AS-DI) values. We define the AS-DI value as the value that the DI-curves approximate when frequencies tend towards zero. Paper D provides the frequency dependent DI-curves and a more thorough assessment of the free-field cases.

The evaluation of our 3D DI surface-area weighting scheme reveals that theoretically correct AS-DI values were obtained for Syn_{1998}, whereas Syn_{1550} showed a small drop in AS-DI. This error can be contributed to the missing source positions in the lower part of the sphere. The AI-DI values show a maximal drop of 0.2 dB for both cases compared to the AS-DI values. These errors are solely caused by the averaging process and by a DI roll-off towards high frequencies. The roll-off is caused by wavelength of the sound becoming comparable

		Front-omni	Cardioid	Optimal
AS-DI	Theoretical	0.0 dB	4.8 dB	6.0 dB
	Syn_{1998}	0.0 dB	4.8 dB	6.0 dB
	Syn_{1550}	0.0 dB	4.6 dB	5.6 dB
AI-DI	Syn_{1998}	0.0 dB	4.6 dB	5.9 dB
	Syn_{1550}	0.0 dB	4.5 dB	5.6 dB
	Unw_{1550}	0.0 dB	3.6 dB	4.4 dB

Table 10.1: AI-DI and asymptotical DI (AS-DI) values. Theoretical values are obtained from [Dil12], whereas the remaining values are calculated based on synthesized free-field responses for two microphones placed 10 mm apart. Unw_{1550} show the unweighted Syn_{1550}.

to the the microphone distance, see Paper D, and cannot be contributed to our weighting scheme.

The unweighted AI-DI values, Unw_{1550}, show a dramatic drop for the cardioid and optimal filters of 0.9 and 1.2 dB compared to the Syn_{1550}. This indicates that our DI weighting scheme, to a great extend, compensates for missing source positions, even when source positions are missing on a large region of the sphere.

10.5 BTE optimization of directional filters

We optimize directional filters for a BTE hearing aid using the weighting method and optimization procedure described in Section 10.2 and 10.3.

Four kinds of optimized directional filters are calculated. The first kind is directional filters optimized on synthesized HRTFs for a sphere. The second kind is directional filters optimized for FEM simulated HRTFs (Chapter 7) for a head model of subject NH167. The third kind is directional filters optimized on HRTFs measured on the Brüel & Kjær HATS head and torso simulator using the Kongebakken measurement setup. The last kind is directional filters optimized on HRTFs measured for the human subject NH167 using the ARI measurement setup. The four kinds of optimized directional filters will be referred to as *sphere optimized*, *simulated*, *HATS optimized* and *human*. The simulated and human directional filters are individualized and they are therefore optimized for the left and right ear separately. The HATS optimized filters are however non-individualized and there is no reason for assuming that the left-ear HATS filter

would perform better on our human subjects left ear than the right-ear HATS filter. Therefore a single HATS filter is obtained, which is an average over the left and right ear filters.

10.5.1 Sensitivity patterns

The behavior of our four kinds of directional filters can be evaluated by their sensitivity patterns in free field. Free-field responses were synthesized in Section 9.1.1. It is known that the optimal directional filter in free field is a hyper-cardioid [Dil12], and the sensitivity pattern of such a filter was shown for a single frequency in Figure 3.4. Directional filters are however frequency dependent. The first column of plots in Figure 10.1, shows the sensitivity patterns for a hyper-cardioid in free field for 0.25, 0.5, 1, 2, 4, 8 kHz. The sensitivity patterns are very similar across the shown frequencies.

The behavior of a directional microphone is very well understood in free field, however when a simple sphere (imitating a head) is introduced the situation becomes more complicated. We calculate optimal filters based on the spherical model explained in Section 9.1.2. The sensitivity patterns are seen in the second column of Figure 10.1. The sensitivity patterns become more frequency dependent when the sphere is introduced, because reflections and diffractions on the surface of the sphere change the sound field. The sensitivity patterns look similar to the hyper-cardioid, especially at high frequencies.

Column three to five in Figure 10.1 show sensitivity patterns for the simulated, HATS and human optimal filters. The sensitivity patterns for all three cases become more complicated because pinna and torso are introduced. All three filters have sensitivity patterns at 0.25 kHz that approach an omni-directional pattern. This behavior can be explained, for the directional filters obtained from measured HRTFs, by a microphone mismatch. A microphone mismatch would mostly influence measurements at low frequencies.

The sensitivity patterns obtained for the human optimized directional microphone, Figure 10.1 column five, are very similar to the patterns obtain for the sphere optimized filters. The sensitivity patterns for the simulated and HATS optimized filters, in column three and four, are, on the contrary, very different from sensitivity patterns for the sphere at some frequencies. The simulated filters yield sensitivity patterns that resemble a cardioid at low frequencies, whereas the sensitivity patterns are more similar to the sphere at higher frequencies. The sensitivity patterns for the HATS optimized filters resemble the sphere's sensitivity patterns to some extend, however the directional nulls are shifted forward.

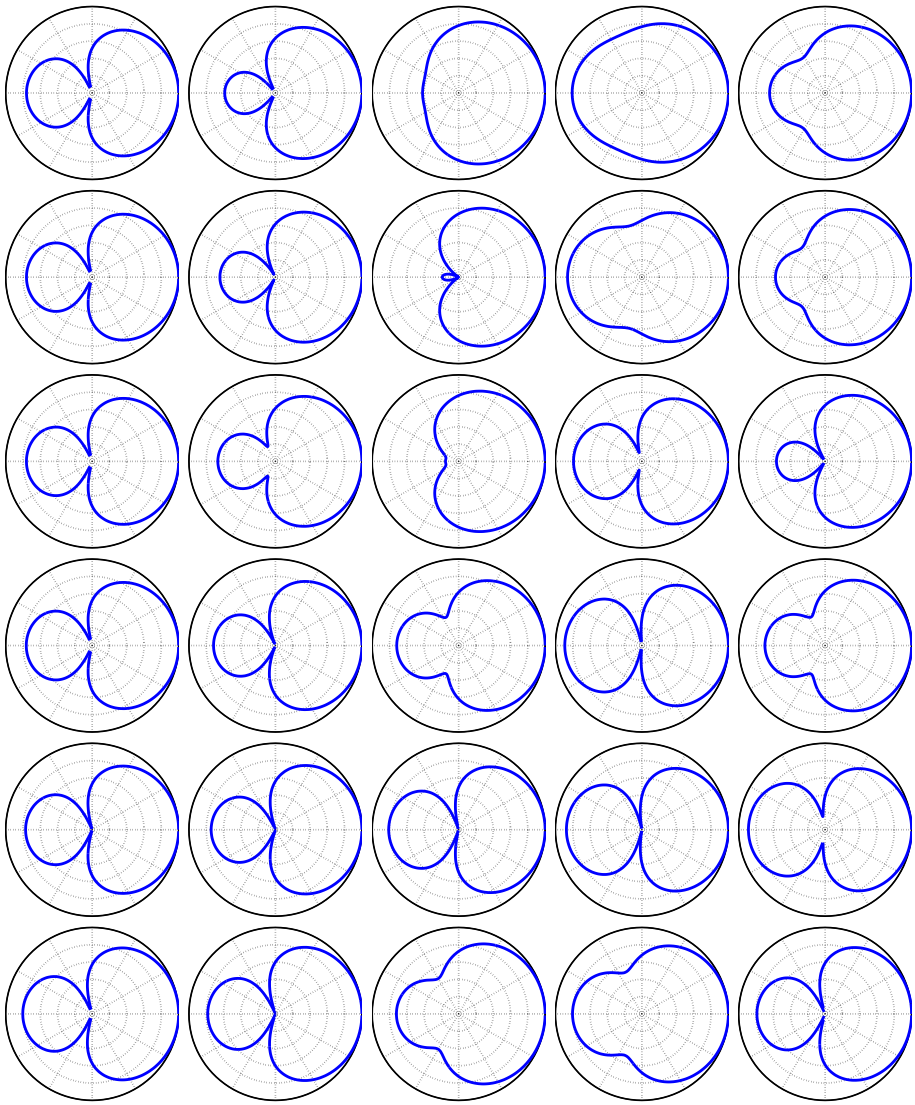


Figure 10.1: Sensitivity patterns in free field. Axes have been removed for better visualization. Zero degree azimuth is to the right and axes are equal to those showed in Figure 3.3. Rows from top to bottom: 0.25, 0.5, 1, 2, 4, 8 kHz. Columns: Hyper-cardioid, sphere, simulated, HATS and human optimized directional filters.

The behavior of the sensitivity patterns for the simulated and measured HRTFs are difficult to interpret, since the presence of pinna and torso complicates the sound field for which the microphones were optimized. The exact behavior of the sensitivity patterns for the three directional filters are therefore noted without an attempt to explain the behavior further.

10.5.2 DI performance

An evaluation of the DI performance of our optimized directional filters in free field are not particularly interesting, since the filters are optimized on either a head model, a printed head or a human head. It is however interesting to visualize the filter performances on a sphere, because it provides a more 'clean' result, which can be compared to the more complicated behavior obtained for measured or simulated data. Figure 10.2 shows DI-curves for optimal, front-omni and hyper-cardioid directional filters applied on the sound field of a sphere. All curves appear relatively constant along the frequency axis. The front-omni curve is located approximately at 0 dB. The deviations from the 0 dB line is caused by the presence of the sphere. The optimal curve is, as expected, located around 6 dB, along with the hyper-cardioid.

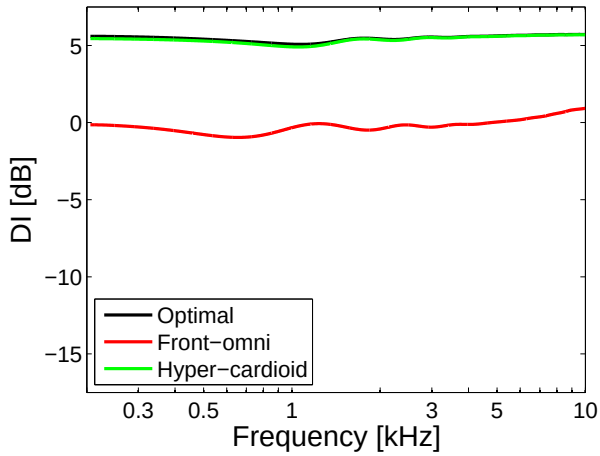


Figure 10.2: DI-curves for synthesized spherical head-model HRTFs. The figure shows the results obtained with directional filters: Optimal, front-omni and hyper-cardioid.

Recall that the final goal of this thesis is to evaluate the two verification steps that were introduced in Figure 1.2. Figure 10.3 provides the results obtained for

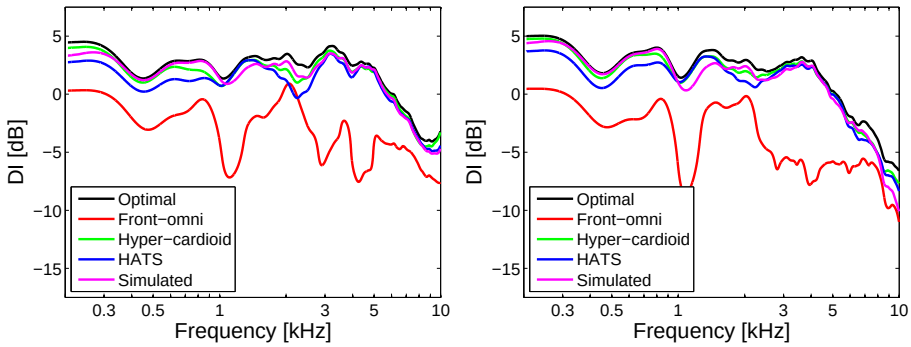
this evaluation. The figure shows DI-curves calculated for test subject NH167 and NH169, which is the printed version of subject NH167's head model. Figure 10.3 a and b show the final verification on the human test subject, whereas c and d contains the intermediate verification using the printed head model.

One thing to notice is that the DI-curves for a human head do not correspond to those obtained for a sphere. The DI-curves for the front-omni filters have many dips and peaks, and the values are often less than 0 dB, which means that the sensitivity towards the sound arriving from the front is less than the average sensitivity from all directions. DI-values below 0 dB can be explained by the BTE position behind pinna, because pinna has a shadowing effect for the sounds arriving from the front. Pinna effects does however mostly have an effect at high frequencies. We believe that reflections from torso and thighs are responsible for the large dips at 1 and 3 kHz, and that they, along with reflections and diffractions from the head, contribute with the much fluctuating front-omni DI-curves.

Optimal, front-omni, hyper-cardioid, HATS-optimized, and simulated directional filters were applied to the measured HRTFs and DI-curves were calculated using the proposed weighting method. Table 10.2 provides AI-DI values. The final verification shows that the simulated filters provide DI-values that are comparable to the optimal DI-curve. The reduction in AI-DI values from simulated to optimal filters are 0.5 dB, for both measurements on subject NH167 (NH167 and NH167-2). This is a relatively low value, since an additional 1 dB of AI-DI corresponds to about 10 % better speech intelligibility in critical listening conditions, according to Dillon [Dil12]. The hyper-cardioid and HATS optimized filters produce AI-DI values in the same range as the simulated filters, which means that the improvement from changing to individualized directional filters is quite small for test subject NH167.

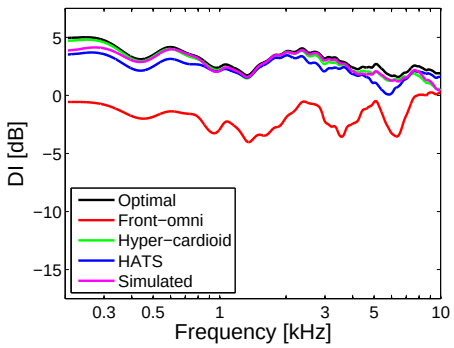
The intermediate verification reveals that the simulated filters perform even better on the printed head than on the human subject. This suggests that the simulation has better agreement with HRTFs measured on the printed head than with measurements on the human subject. The higher agreement could arise due to more similar geometries (virtual model vs. human head/printed model) or more similar acoustic properties of material (virtual material vs. human skin/printing material).

Figure 10.4 shows DI-curves for the remaining three test subjects and Table 10.2 provides AI-DI values for all individuals. The results show that the non-individual filters perform very well for some individuals and worse for others. In particular subject NH168 and NH170 left are have a several dB improvement from the non-individual DI-values to the optimal values for some frequencies. The most critical frequencies are approximately 2 and 6 kHz for NH168 and

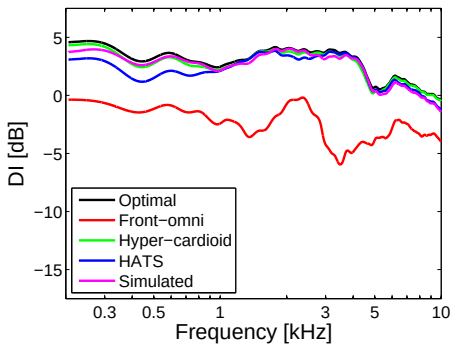


(a) Subject 167 left ear.

(b) Subject 167 right ear.



(c) Subject 169 left ear.



(d) Subject 169 right ear.

Figure 10.3: BTE hearing aid: DI-curves for subject NH167 and NH169 (printed head). The figure shows the results obtained with directional filters: Optimal, front-omni, hyper-cardioid, HATS optimized and simulated.

			Relative to F.O.			
	F.O.	Opt.	Opt.	H.C.	HATS	Sim.
NH166 L	-2.1	3.3	5.4	5.1	4.7	-
NH166 R	-1.8	3.1	4.9	4.3	4.0	-
NH167 L	-2.0	2.8	4.8	4.2	3.7	4.3
NH167 R	-2.4	2.8	5.2	4.7	4.3	-
NH167-2 L	-2.4	2.6	5.0	4.6	4.2	4.5
NH167-2 R	-2.1	3.5	5.6	5.3	5.0	-
NH168 L	-1.1	3.6	4.7	4.0	3.6	-
NH168 R	-2.0	3.1	5.1	4.4	3.9	-
NH170 L	-2.9	3.2	6.1	5.0	4.9	-
NH170 R	-1.3	4.1	5.4	5.1	4.8	-
NH169 L	-1.9	3.2	5.1	4.8	4.4	4.8
NH169 R	-2.0	3.3	5.3	5.1	4.7	-
Mean	-2.0	3.2	5.2	4.7	4.4	-
SD	0.5	0.4	0.4	0.4	0.5	-

Table 10.2: BTE hearing aid: AI-DI values calculated for different directional filters based on measured HRTFs. Optimal (Opt.), hyper-cardioid (H.C.), HATS, and simulated (Sim.) values are given relative to front-omni. Mean and standard deviations (SD) are provided except for the simulated filters.

2-3 kHz for NH170 left. The high non-individualized AI-DI values for subject NH167 leaves very little room for improving the AI-DI using simulated filters.

Even though the benefit of directional filters optimized on simulated HRTFs was limited for test subject 167 other individuals might benefit more from such individual filters. The HATS filters, that we believe are somewhat similar to the filters used in commercial hearing aids, are measured for a head and torso simulator that represents an average of the human adult population. Individuals who differ from the average population might benefit more from individual filters. Two groups of individuals who are interesting in this regard is children, which have radically different head sizes and proportions than an average adult, and elderly people, who have larger ears because the ears continue to grow throughout life. Both groups could possibly benefit more from listener-specific directional filters.

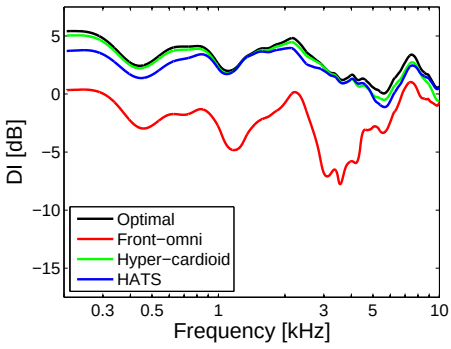
10.5.3 A child-sized sphere

It is interesting to investigate the impact of individual directional filters for children compared to the normal filters. However, HRTF measurements on a child is not feasible because it requires that the child sits still for a long period of time. It is however easy to synthesize spherical head-model HRTFs for a child-sized sphere. Here a sphere with a diameter of 12 cm has been used. Figure 10.5 shows DI-curves for the 12 cm sphere using optimal, front-omni, hyper-cardioid, and adult (diameter of 17.5 cm) optimized directional filters. The figure shows that the benefit from using directional filters optimized on a child-sized sphere instead of an adult-sized sphere is less than 0.1 dB.

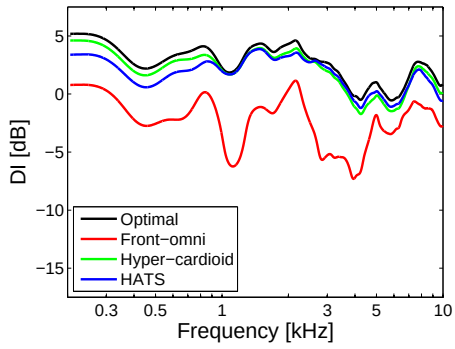
Even though it is tempting to conclude that a child would not benefit much from having directional filters optimized for a child instead of an adult, one has to be careful with such conclusions because the sound field surrounding the head is more complicated than for a sphere, due the reflections and diffractions on torso and pinna.

10.6 ITE optimization of directional filters

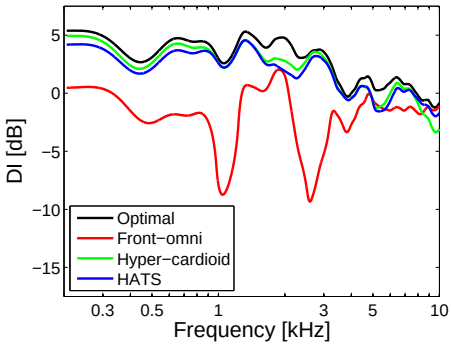
In addition to the DI-optimization of directional filters in a BTE hearing aid, we also optimize directional filters for an ITE hearing aid. The weighting method and optimization procedure described in Section 10.2 and 10.3 are used again. For the ITE hearing aid only HRTFs measured on human test subjects using the



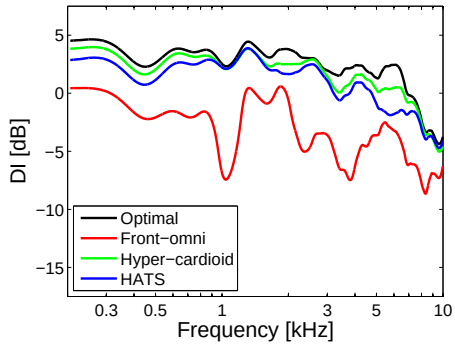
(a) Subject 166 left ear.



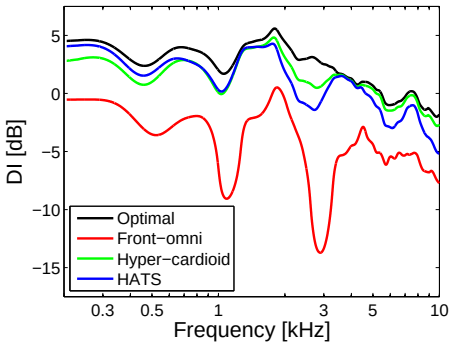
(b) Subject 166 right ear.



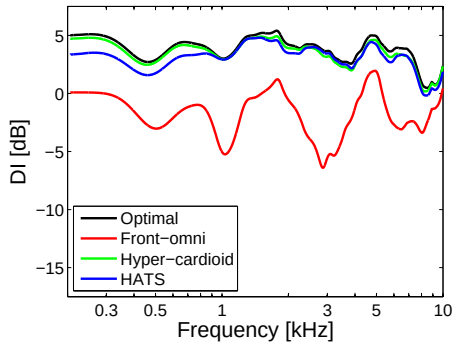
(c) Subject 168 left ear.



(d) Subject 168 right ear.



(e) Subject 170 left ear.



(f) Subject 170 right ear.

Figure 10.4: BTE hearing aid: DI-curves for subject NH166, NH168 and NH170. The figure shows the results obtained with directional filters: Optimal, front-omni, hyper-cardioid and HATS optimized.

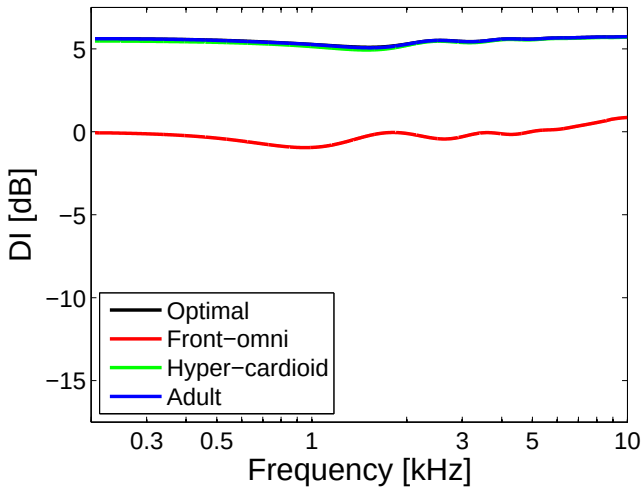


Figure 10.5: DI-curves for child-sized (12 cm) spherical head-model HRTFs. The figure shows the results obtained with directional filters: Optimal, front-omni, hyper-cardioid, and adult (17.5cm) optimized.

ARI setup are available. No simulations were made using an ITE microphone placement, and HRTFs measured on a printed head or on a HATS were not available either. Furthermore, HRTFs were only measured for three human test subjects (NH166, NH167 and NH168), because ITE shells were not manufactured for test subject 170. The three HRTF measurements were supplemented by a remeasurement on test subject NH167, referred to as NH167-2.

Based on the HRTFs measured for the three human test subjects using an ITE hearing aid and the ARI measurement setup nine directional filters were calculated. Optimal directional filters were calculated for each individual for each ear, leading to six individual filters, plus an additional two directional filters optimized for the remeasurement. Finally, an average of the six initial filters were calculated, leaving out the remeasurement. The remeasurement, NH167-2, of subject NH167 was excluded from that was optimal for an average population. Because our set of individuals consist of only three subjects, it is far from an average of the population. However, the average filter can provide an indication of how well an average filter would perform. The filters optimized on measured HRTFs for the test subjects are referred to as *human* and the average filter is referred to as *average*.

10.6.1 Sensitivity patterns

Figure 10.6 shows sensitivity patterns for directional filters optimized on measured HRTFs for an ITE hearing aid placed on the right ear of the three human subjects and for the human average directional filter. The directional filters vary across individuals, however compared to the directional filters optimized on HRTFs measured for a BTE hearing aid on a human subject (Figure 10.1 column five) the ITE directional filters all have directional nulls which have been shifted forward. This change in the sensitivity pattern could be caused by the effect of the pinna which is present for the ITE placed in the ear canal. The inter-individual differences might be caused by differences in pinna shape, and in particular by differences in concha shape, see Figure 2.2, because sounds arriving at the ear canal have been subjected to reflections and diffractions in the concha.

10.6.2 DI performance

Figure 10.7 and 10.8 show DI-curves calculated from HRTFs measured on ITE hearing aids for subject NH167 and NH167-2, and NH166 and NH168, respectively. The DI-curves are calculated based on optimal, front-omni, hypercardioid, and human average directional filters.

The front-omni directional filters for the ITE hearing aid result in higher DI-values generally, than what was obtained for the BTE hearing in Figure 10.3 and 10.4. The lower front-omni values obtained for the BTE hearing aid arise due to the microphone placement behind the ear. The ear contributes with a shadowing effect for the sounds arriving from the front resulting in low DI-values for the front-omni filters. The ITE hearing aid is on the contrary positioned in the ear canal and the effect of pinna is thereby preserved. The pinna has a shadowing effect for sounds arriving from the rear, resulting in higher DI-values for the front-omni filters for the ITE than for the BTE. Table 10.3 provides frequency-averaged AI-DI values. For the ITE hearing aid we obtain AI-DI values in the 0 dB range for the front omni alone. The optimal effect of pinna is obtained for sounds arriving approximately from a 40° azimuth angle and not from the front. This was observed in Figure 9.9, and is most likely the reason for the AI-DI values of approximately 0 dB instead of higher AI-DI values. We show a mean front-omni AI-DI value of 0.23 dB with a standard deviation of 0.27 dB for the ITE hearing aid and a mean front-omni value of -2.0 dB with a standard deviation of 0.5 dB for the BTE hearing aid. Similar values were reported by Ricketts and Mueller [RM99].

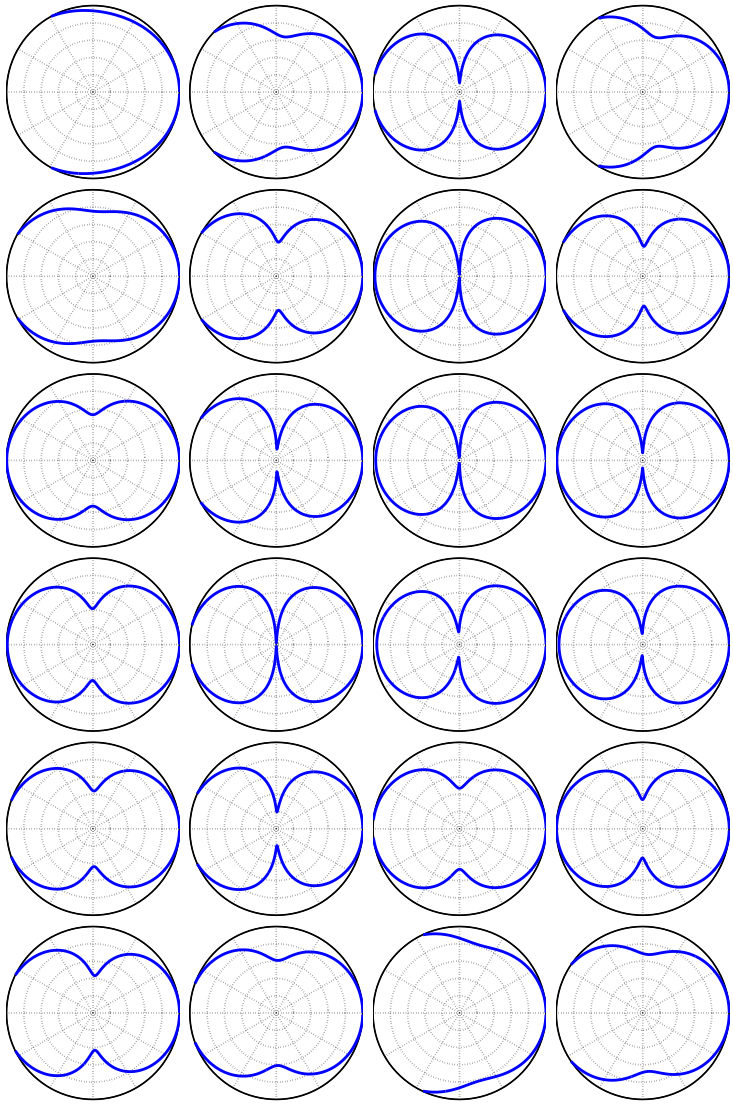
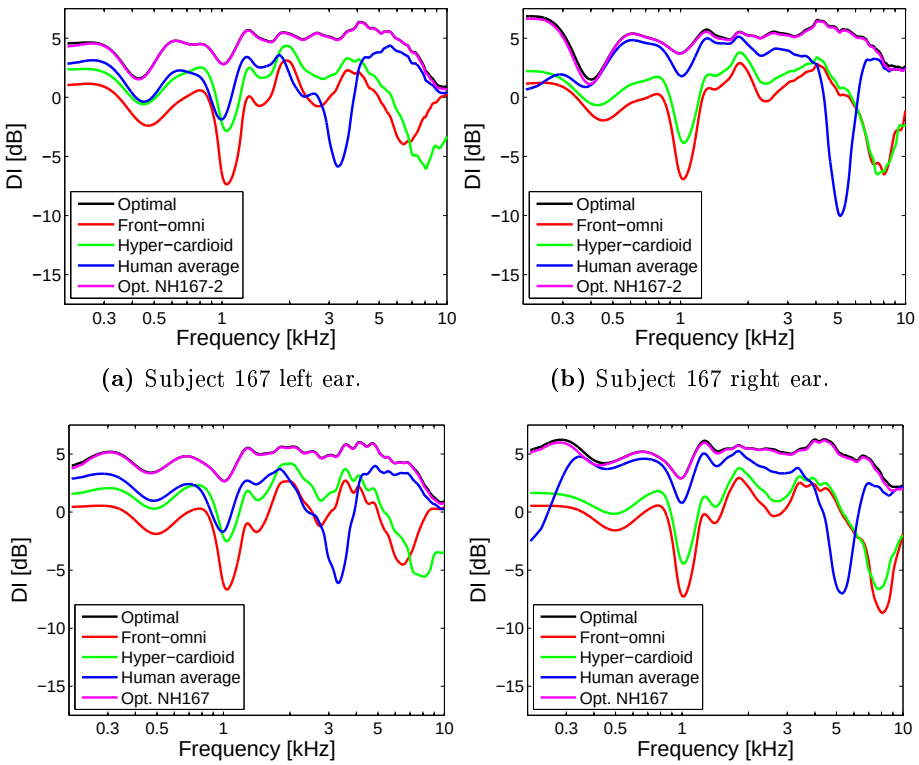


Figure 10.6: Sensitivity patterns in free field. Axes have been removed for better visualization. Zero degree azimuth is to the right and axes are equal to those showed in Figure 3.3. Rows from top to bottom: 0.25, 0.5, 1, 2, 4, 8 kHz. Columns: NH166 right ear, NH167 right ear, NH168 right ear optimized and average directional filters.



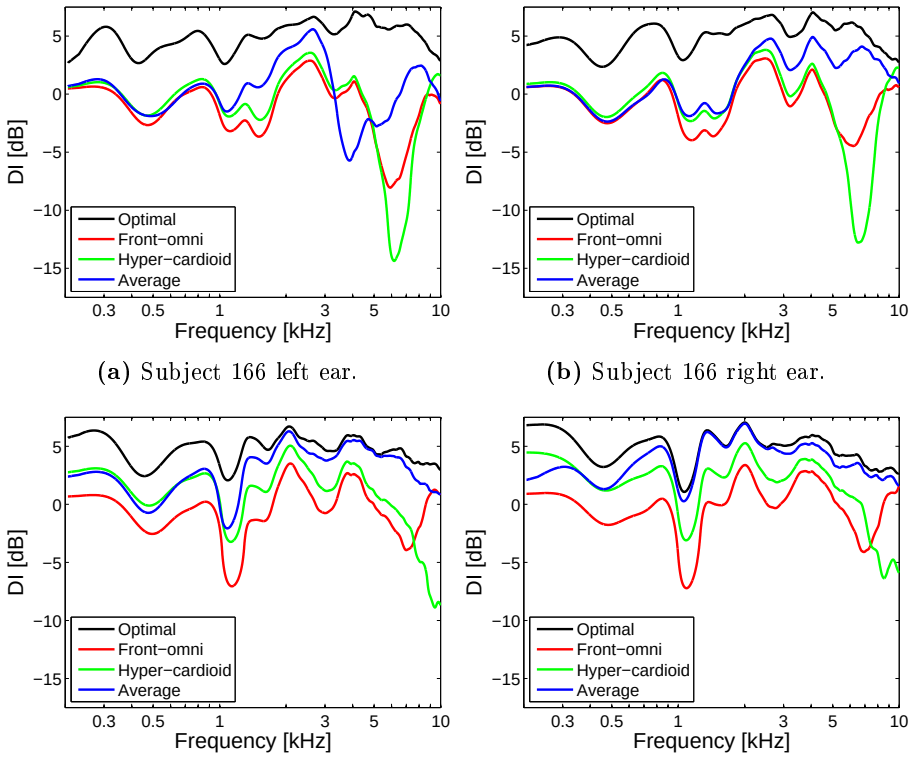
(a) Subject 167 left ear.

(b) Subject 167 right ear.

(c) Subject 167-2 left ear.

(d) Subject 167-2 right ear.

Figure 10.7: ITE hearing aid: DI-curves for subject NH167 and the remeasurement NH167-2. The figure shows the results obtained with directional filters: Optimal, front-omni, hyper-cardioid, human average, and optimal filters of subject NH167/NH167-2 for cross validation.



(a) Subject 166 left ear.

(b) Subject 166 right ear.

(c) Subject 168 left ear.

(d) Subject 168 right ear.

Figure 10.8: ITE hearing aid: DI-curves for subject NH166 and NH168. The figure shows the results obtained with directional filters: Optimal, front-omni, hyper-cardioid, and human average.

			Relative to F.O.		
	F.O.	Opt.	Opt.	H.C.	H.A.
NH166 L	-0.27	5.2	5.5	0.69	1.6
NH166 R	-0.013	5.4	5.4	0.65	1.8
NH167 L	0.21	4.8	4.6	1.7	1.6
NH167 R	0.46	5.1	4.6	1.0	3.0
NH167-2 L	0.22	4.8	4.6	1.7	1.6
NH167-2 R	0.44	5.2	4.8	0.97	3.1
NH168 L	0.25	5.1	4.8	1.8	3.4
NH168 R	0.54	5.2	4.7	2.1	4.0
Mean	0.23	5.1	4.9	1.3	2.5
SD	0.27	0.2	0.4	0.6	1.0

Table 10.3: ITE hearing aid: AI-DI values calculated for different directional filters based on measured HRTFs. Optimal (Opt.), hyper-cardioid (H.C.), and human average (H.A.) values are given relative to front-omni. Mean and standard deviations (SD) are provided.

Furthermore, it is noticed that the DI-curves are significantly lower using hyper-cardioid filters on the HRTFs measured for the ITE hearing aids, than for the HRTFs measured for the BTE hearing aids. The hyper-cardioid filters are the optimal directional filters in free field, and the ITE placement in the ear canal probably results in a sound field which is further from a free sound field than what is obtained using a BTE hearing aid placed on top of the ear.

The human average directional filters result in improved DI-values compared to the hyper-cardioid filters. However the DI-values obtained with the human average filters are still much lower than the DI-values obtained for the optimal filters. AI-DI values are seen in Table 10.3. The mean improvement in AI-DI values from human average to optimal filters is 2.4 dB, with a maximum improvement of 3.6 dB, corresponding to about 36% improvement in speech intelligibility. An improvement in speech intelligibility of 36% would definitely be a worthwhile improvement for the hearing-aid user.

The DI-curves for front-omni, hyper-cardioid, and human average directional filters all contain large dips in the curves. The front-omni curves for the three test subjects all have dips around 1 kHz and 7-8 kHz. The 1 kHz dip is most likely caused by torso reflections, whereas the 7-8 kHz dip is probably caused by reflections and diffractions in the concha. The hyper-cardioid and human average filters contain large dips, which are more varying among individuals, placed at frequencies above 3 kHz. These dips are probably caused by the complicated

sound field in the concha, which the non-individualized filters cannot account for.

Figure 10.7 also shows a cross validation of the optimized directional filters (magenta colored curves) for subject NH167 and the remeasurement NH167-2. The cross validation reveals that the directional filters, optimized on HRTFs measured on two different days with a repositioning of the ITE hearing aid, performs equally well on both sets of HRTFs. It therefore appears that the individual shape of concha are highly important for the optimization of directional filters of an ITE hearing aid and that the re-positioning of the ITE hearing aid is relatively robust, since robust directional filters were obtained.

Considering the large room for improvement of the directional filters in an ITE hearing aid, it would be very interesting to generate individual simulated filters for an ITE hearing aid based on a listener-specific head model. The framework for simulating HRTFs for optimization of the directional microphone in an ITE hearing is already available, since the method explained in this thesis for at BTE hearing aid can be applied directly. An accurate head and ear model is, however, required. The head models, which we presented in Chapter 4, were shown to have difficulties with occluded areas in the pinna region. Because the shape of concha appears to be highly important for optimization of the directional microphone in an ITE hearing aid, one might consider a supplement ear impression for improving the shape of concha in the head model. A 3D scan of the concha impression can be stitched together with the head model using the framework presented in Chapter 4. ITE hearing aids are already manufactured based on an ear-impression and this already existing ear impression could be extended to include concha. A 3D scan of the ear impression would provide a more accurate model of concha, probably leading to more accurate HRTF simulations for the ITE hearing aid and thereby more accurate individual simulated directional filters.

Chapter 11

Conclusions

The overall goal of this thesis is to improve speech intelligibility for hearing-aid users by individualizing the directional microphone. The idea is to facilitate easy acquisition of individually optimized directional filters through a three-step pipeline. The first step is to estimate a listener-specific 3D head model based on 2D images, the second step is to simulate individual HRTFs based on the estimated 3D head model, and the final step is to calculate optimal directional filters based on the simulated HRTFs.

We present a framework for directional microphone optimization using HRTFs simulated from a listener-specific head model. The framework consists of generation of a listener-specific 3D head model, simulation of listener-specific HRTFs based on the 3D head model, and calculation of optimal directional filters based on the simulated listener-specific HRTFs. The framework was applied to a BTE hearing aid.

The directional filters in a BTE hearing aid, optimized from simulated listener-specific HRTFs, provide a directionality which is comparable to the optimal directionality obtained from measurements on the human subject. The reduction in AI-DI from optimal to simulated filters is 0.5 dB for the BTE hearing aid, which corresponds to 5% less speech intelligibility. A comparison against non-individual directional filters (hyper-cardioid and HATS optimized) revealed equally high AI-DI values for test subject NH167. However, looking at our other

test subjects, it is noticed that the non-individual filters perform very well for some individuals and worse for others. At some frequencies, several dB improvement is seen from the non-individual DI-value to the optimal value. The high DI (and AI-DI) values for subject NH167 leaves very little room for improving the AI-DI using individual filters.

We developed a pipeline for 3D printing of a compete listener-specific human head. The 3D printed head facilitated an extra verification step of the simulated directional filters. We showed that the difference between the AI-DI performance of optimal and simulated directional filters were smaller for the printed head than for the human subject. A better agreement between simulation and printed head is not surprising since the printed head excludes some of the differences that exist between HRTFs measured on a human and HRTFs simulated on a model. The differences that are excluded are geometrical differences in the head and torso geometry, differences in acoustical properties, and disturbances from human movements. The higher agreement between printed head and simulation suggests that the performance of the simulated filters could be improved further if these issues were addressed. On the other hand, the optimized simulated filters are currently based on a head model obtained through our highly controlled scanning, stitching and meshing process and the obtained head model is therefore of very high quality. Our initial goal was to estimate 3D head models from 2D images and a statistical shape prior. This step is still a work in progress, because both the process of fitting a 3D statistical model to 2D images, and the generation of a 3D statistical head model turned out to be more complicated than initially believed. Because the 2D image approach is less controlled than our scanning procedure, it cannot be assumed that the resulting head models have the same accuracy.

It was shown that the benefit from directional microphones using non individualized filters varies across the population [VFP95], and some individuals will therefore have a larger room for improvement from individualized filters. Because standard directional filters are optimized to achieve the highest AI-DI average across the population, people who deviate from the average population might have a larger benefit from an individualized directional microphone. Children and elderly, who differ from the average, are therefore of interest. Children have a smaller head size and smaller ears than average, and elderly individuals have larger ears (pinnas), because the ears continue to grow throughout life. High quality hearing aids are particularly important for small children who are learning speech. However, producing good hearing aids for the elderly is also important because a large fraction of hearing-aid users consists of elderly individuals.

The BTE hearing aids show very little room for improvement using individualized directional filters, even though some individuals might benefit more from

individualized directional filters. We showed that the directional filters in ITE hearing aids had an improvement in AI-DI values of up to 3.6 dB between an average filter and an optimal filter. We believe that the large benefit from using individualized filters for an ITE hearing aid arise because the individual filters can account for the complicated and very individual sound field in the concha. This suggests that hearing-aid users with ITE hearing aids potentially could have a large benefit from individualized directional filters.

Considering the large room for improvement of the directional filters in an ITE hearing aid, it would be very interesting to test the proposed pipeline for generating individual simulated filters on an ITE hearing aid. The head models which was generated for this thesis can be employed, however to obtain a very accurate 3D model of concha we suggest that the 3D head model is supplemented with an ear impression. The ear impression which is already made for ITE hearing-aid manufacturing could be extended to include concha, and a 3D scan of the ear impression would provide an accurate concha model. HRTF simulations on the head and concha model, and subsequent optimization of directional filters using the simulated HRTFs would probably lead to accurate individual directional filters. We believe that these individual directional filters will lead to an improved intelligibility for ITE hearing-aid users.

Appendix A

A three dimensional children head database for acoustical research and development

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 1pPPa: Binaural Hearing and Binaural Techniques I

1pPPa5. A three dimensional children head database for acoustical research and development

Stine Harder*, Rasmus R. Paulsen, Martin Larsen and Søren Laugesen

***Corresponding author's address: Compute, Technical University of Denmark, Richard Petersens Plads, Lyngby, 2800, Denmark, sthar@imm.dtu.dk**

Most computational-acoustic work within spatial hearing relies on head-related transfer functions from databases of measurements taken on adult humans or dummy heads. We aim to provide a set of 3D digital heads including children, from which head-related transfer functions can be computed instead of measured. However, current volumetric scanning techniques do not have sufficient resolution for accurately scanning the external ear, and computed tomography also involves radiation. In this paper we propose a framework for scanning, stitching and meshing complete human heads. The process starts by acquisition of multiple 3D surface scans of the same subject using a high-resolution photogrammetric scanner. Secondly, the scans are semi-automatically aligned and noise and incoherence is removed. This is followed by an iterative process where a volumetric implicit representation of the head is optimized. The process consists of a regularized surface-reconstruction step followed by an alignment step. Finally, a surface representation of the entire head is extracted using a triangulation of the zero-level iso-surface of the implicit volume. The process has been used to reconstruct the heads of children aged 10 months to 9 years. The data and the associated reconstruction algorithms will be made publicly available for use in acoustical research and development.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800159]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Spatial hearing in general and binaural hearing in particular depend on physical spatial acoustic cues. These acoustic cues are traditionally divided into monaural cues, which comprise the direction-dependent acoustic filtering of the pinna, head, and torso, and the binaural cues, which occur because humans have two ears, one on each side of the head. For modeling purposes, both types of cues may be captured in the Head-Related Transfer Functions (HRTFs). The HRTF represents the transfer function from a far-field sound source to a sound pressure receiver at the entrance of the blocked ear canal, relative to the the sound pressure at the center of the head in the free field (head absent). HRTFs are ear-specific and depend on the direction of sound incidence [1]. HRTFs can also be determined for near-field sound sources, in which case there is an additional dependence of distance [2].

A mathematical representation of spatial acoustic cues is useful for many purposes: e.g. models of spatial hearing [3], fitting of hearing aids [4, 5], and creating virtual 3D audio over headphones [6, 7]. In some of these applications, generic HRTFs suffice (e.g. most modeling work), while in other applications it is necessary to consider the variation in HRTFs across the population (e.g. when fitting hearing aids to adults as well as children [8]. In the latter case, databases of HRTFs measured across many people are helpful, such as the CIPIC database [9]. However, compiling such databases is very cumbersome in terms of measurements.

The present work is an initial step towards finding an alternative method to capture the HRTFs of an individual. The idea is to replace the cumbersome measurements with HRTF simulations in a computer. HRTF simulations requires accurate 3D digital heads. The SYMARE database by [10] is a database consisting of head-related impulse responses for 60 listeners along with surface mesh models obtained from magnetic resonance imaging (MRI) data. The resolution of the MRI data is however not sufficient for accurately scanning the external ear and the process of obtaining surface mesh models from MRI data is very complex. Furthermore MRI scanning of young children requires sedation. Computed tomography involves radiation and is not an appropriate imaging modality. The present article uses 3D surface scanning and contributes with a framework for scanning, stitching and meshing complete human heads. The results are high-quality surface representations which can be used for HRTF simulations for children as well as adults. Regarding simulation of HRTFs the surface representations are raw data, that can be converted into a more computational appropriate format with the required level of detail.

DATA

Data was obtained during a scan session at Enheden for Psykiatrisk Forskning, Aalborg Psykiatriske Sygehus using a Canfield scientific Vectra M3 surface scanner. The surface scanner and setup are shown in Fig. 1. A custom made hair covering cap was used, due to the diffuse properties of the hair. Surface scans were captured from multiple angles in order to cover the entire head. Because of the complex structure of the ear, this region was given special attention.

The surface scanner has an accuracy of approximately 0.2 millimeter and each scan consists of around 65.000 vertices and 130.000 triangles. Surface textures are also computed by the scanner firmware as seen in Fig. 2. Data for each individual consist of 7-12 surface scans. Fig. 3 shows a zoom in on the ear region, along with the wire frame. A scan protocol was developed subsequent to the scan sessions.

FIGURE 1: Canfield Vectra M3 scanner consisting of three pods each mounted with two Canon SLR Cameras.

FIGURE 2: Example of surface scans. It can be seen that the used hair covering cap was not optimal for all children. In this case the neck part was too loose.

FIGURE 3: Zoom in on the ear region, along with the wire frame.

METHOD

The goal is to combine several surface scans acquired from different angles into one complete surface model of the entire head and ears. The major obstacle is the shape variation among the different scans. While one acquisition is near instantaneous (six standard cameras triggered simultaneously) it is impossible to avoid changing facial expressions and rotations of the head between the scans. In particular young children can not be expected to stay in the same position during the entire acquisition protocol. The consequence is that the final model should be assembled of a number of sub-parts, where the number of scans used for each sub-part is dependent on the consistency of the sub-part. Consistency is here defined as a sub-part (nose, ear, eye for example) which is representing the true anatomy and has little variation between the set of scans. A good example of a consistent part is the pinna that normally does not deform even with changing head pose and facial expression. The mouth and neck regions normally vary significantly between scans.

While it theoretically would be possible to develop a completely automated registration and

merging algorithm, we have chosen to use a semi-automated approach to create a small but high-quality and controlled dataset. This dataset can potentially be used as a basis for a learning-based reconstruction algorithm based on for example a strong statistical prior as demonstrated in the widely popular statistical shape models [11].

Manual Annotation

The data is initially manually annotated to mark the areas that are assumed consistent across the scans. Each surface scan was manually marked using the software tool Sumatra [12]. Six anatomical areas were marked with scalar values: face (1), left ear (2), right ear (3), top of the skull (4), bottom of the skull (5) and neck (6). Remaining areas were marked as "unmarked" (0). An annotation of the anatomical areas on a surface scan are shown in Fig. 4. Furthermore, we have defined 93 anatomical landmark points on the face and ear regions, see Fig. 5. The landmarks were placed manually on each surface scan using MeshLab [13]. If a sub-scan does not contain one or several anatomical areas where landmarks should have been placed, the corresponding landmarks are marked as inactive for this scan.

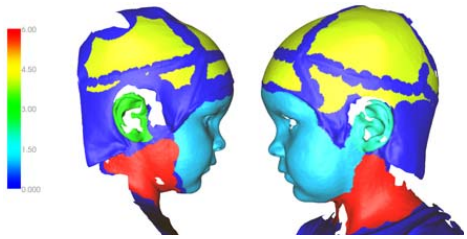


FIGURE 4: Marking of anatomical areas. Blue regions are defined as not marked. It can be seen that areas where the cap was folded and where there is hair has been excluded.

FIGURE 5: 93 predefined landmarks.

Alignment and Reconstruction

One of the scans in a scan set is defined as the reference. This is usually a frontal scan with nearly all landmarks present. The remaining scans are subsequently rigidly aligned to the reference scan by determining for each scan the set of landmarks that are defined in both the current scan and the reference scan. The rigid-body transformation (3 rotations and 3 translation parameters) is determined using a closed-form solution with quarternions [14]. The relevant marked areas are also extracted from each aligned mesh. The result is a collection of roughly aligned point clouds as seen in Fig. 6.

It is clearly seen on the roughly aligned point clouds that parts of the head is well covered by samples but several regions clearly lack well defined sample points. Furthermore, the scans are

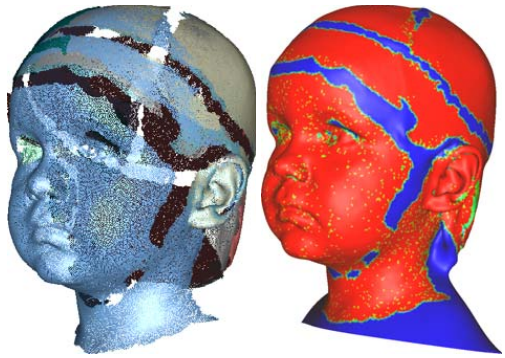


FIGURE 6: Left: Roughly aligned point clouds. Right: The resulting surface visualizing the confidence map where red is high confidence and blue no confidence.

not perfectly aligned and both Gaussian noise and outliers are present. To be able to reconstruct a full head model we therefore need a surface reconstruction and alignment algorithm that is 1) robust to noise, 2) can interpolate in areas with missing data in an anatomically plausible way, and 3) can accurately align the partially overlapping scans. Most surface reconstruction algorithms can handle a certain amount of noise and can close holes. Hole closing normally employs a mathematical or physical assumption on the behavior of the surface in ares with no samples. A classical approach is based on the Delanauy triangulation [15] that will create linear patches covering holes. A physical approach is that the membrane energy of the surface in the hole is minimized, mimicking a soap bubble [16, 17]. Other methods are based on variational approaches as for example regularization of the mean curvature of the surface [18]. We are basing our approach on a higher order mathematical model introduced in [17] that has been shown to be very well suited for human head scans. Inspired by [19] the method proposed consists of several steps that are iterated until convergence. Initially, a single oriented point set with consistent normal directions is created from the input data. Secondly, a signed distance field is computed based on the oriented point set. In the next step a Markov Random Field based regularization is applied to the distance field. In the last step, the input point sets are aligned to the implicit zero level using an implicit variant of the iterative closest point algorithm (ICP) [20]. These steps are then repeated in a multi-scale framework. Finally, the resulting surface is extracted from the regularized distance field using an iso-surface extraction algorithm, and the resulting mesh is optimized. Conceptually, the method keeps track of an ideal implicit surface representation of the combined input data. This implicit surface representation is updated and refined until it optimally represents the aligned input point sets.

Step 1: Point set Merging In the first step, the input point sets are merged into one point set. At this point, they are only roughly aligned and the merged set will therefore contain points that are seen as both outliers and noise. In this application we use the point normals provided by the scanner. Alternatively, consistent point normals can be computed using a local principal component analysis followed by a graph-based voting scheme [17].

Step 2: Computing the Signed Distance Field The distance field is represented as a uniform voxel volume, where the bounds of the volume are computed to extend five voxels beyond the bounds of the merged point set. The signed distance is computed in each voxel as the distance from the voxel center to the line spanned by the five closest points and the average of their normals. Optimally, the zero set of this distance field defines the sought after surface.

However, this initial distance field suffers from ambiguities in regions with holes in the merged point set and in regions with noise. In the following, the initial distance field is defined as $\mathbf{d}^o$ and the initial distance at voxel i as d_i^o. We use a simple one dimensional indexing of the voxels instead of the standard (x, y, z) indexing.

Step 3: Markov Random Field Regularization To remove the influence of noise and introduce hole-filling capabilities, the signed distance field is regularized using a Bayesian approach. The distance field, $\mathbf{d}$, is formulated as a Markov Random Field and the goal is to compute the distance field $\hat{\mathbf{d}}$ that maximizes the posterior probability:

$$\hat{\mathbf{d}} = \arg\max_{\mathbf{d}} \; p(\mathbf{d}|\mathbf{d}^o). \tag{1}$$

The Markovianity assumption implies that the involved probabilities can be formulated locally using near voxel neighborhoods. In the following, we are using a 6-neighborhood for each voxel. The $n = 6$ neighbors of voxel i is written as $i \sim j$. The local probabilities are based on the Gibbs distribution and are therefore formulated as energies. The local prior probability is based on differences between neighboring Laplacians:

$$U_{\mathscr{L}}(d_i) = \sum_{i \sim j} (\mathscr{L}(d_i) - \mathscr{L}(d_j))^2, \tag{2}$$

where d_i is the voxel value in the current distance field and $\mathscr{L}(d_i)$ is the Laplacian estimated at voxel i. A discrete approximation of the Laplacian is used $\mathscr{L}(d_i) = \frac{1}{n} \sum_{i \sim j} (d_i - d_j)$. This approximation is known from image processing, where it is normally formulated as a 3x3 kernel. The normalization is chosen to be the number of neighbor voxels and this is constant, except at the borders of the volume. Obviously, this prior favors distance fields with smooth Laplacians. Since the Laplacian of a distance field is proportional to its mean curvature [21], the prior in Eq. (2) favors distance fields with small variations in mean curvature. The prior model can conceptually be understood as how we would like the surface to behave in regions with no samples. However, a model that forces the surface to follow input data is needed and therefore the following observation model is used $U_{\text{obs}}(d_i^o) = (d_i - d_i^o)^2$.

As mentioned earlier, d_i^o is the original distance at voxel i and d_i is the current estimate. Using the Gibbs measure, this energy function leads to $p(d_i^o|d_i) = \exp(-U_{\text{obs}}(d_i^o))$, that is, a Gaussian distribution with mean d_i. The observation model describes the distribution of the observed values given a true underlying distance field. In our case, the initial distance field is the observed data and we are seeking an estimate of the true distance field $\hat{\mathbf{d}}$.

In order to balance the prior and observation model a per voxel confidence measure $\alpha_i \in [0:1]$ is introduced. It is based on the Euclidean distance from the voxel center to the nearest input point d_i^E. Here $\alpha_i = 1 - \min(d_i^E / d_{\max}^E, 1)$, where $d_{\max}^E$ is a user-defined maximum Euclidean distance. A discussion of suitable values of $d_{\max}^E$ can be found in [17]. Confidence values in voxels near input points are therefore one and in regions with no input points zero. In Fig. 6 (right) the pre-computed α-map has been projected into the resulting surfaces and thereby visualizing which part of the resulting surface that is influenced by the prior, the observation, or a combined model. Using Bayes' theorem the prior and the observation terms are combined into a posteriori probability per voxel. Using the Gibbs measure, it becomes:

$$p(d_i|d_j, i \sim j) = \exp\left(-\alpha_i \beta \; U_{\text{obs}}(d_i) - (1 - \alpha_i \beta) \; U_{\mathscr{L}}(d_i)\right), \tag{3}$$

where the global weight β and the local α_i's are used to balance the prior and observation models. Using this, the maximum likelihood estimate of the voxel value d_i given its six neighbors can be computed as a linear combination of neighboring voxel values and the original distance estimate [17]. We wish to compute the distance field that maximizes the a posteriori

conditional probability $\hat{\mathbf{d}} = \arg\max_{\mathbf{d}} p(\mathbf{d}|\mathbf{d}^o)$. The maximization of the posterior probability is transformed into the minimization of the weighted sum of the energy functions for the prior distribution and the observation model over the entire field. The global optimum can be found using the Iterative Conditional Modes (ICM) algorithm [22], where each site is iteratively assigned its local maximum likelihood estimate. Due to the uniform space division, a multiscale ICM solver is fairly easy to implement. In [17] a discussion of alternative solvers can be found. The result of the regularization is a distance field where the behavior of the zero-level iso-surface is well defined, even in areas with no input points. In areas with dense sampling, the iso-surface adheres implicitly to the observation energy and in areas with no input points it follows the prior energy. In the first two or three iterations, a smooth distance field is computed by selecting a low β value. In later iterations, a high β value is used putting more weight on the observation model.

Step 4: Realignment In this step the individual point sets are realigned to the zero-level iso-surface. Since this iso-surface is both smoothed and regularized, the influence of the individual point sets have been blended together. A specialized version of the iterative closest point algorithm (ICP) [20] is used. Each input set is individually realigned to the zero-level iso-surface. For each point in the input point set the closest point on the iso-surface is found. Since we have an approximate distance field this can be done using a Runge-Kutta like gradient descent. The sought after rigid-body realignment transform can now be computed using the closed form solution found in [14]. Finally, a new merged point set is created after the transformation of the input point sets.

Step 1-4 is repeated in a multi-scale approach. The scale is determined by the voxel size used in the distance field representation. A resulting iso-surfaces can be seen in Fig. 6 (right). In practise, the iso-surfaces are not triangulated during the iterations.

Step 5: Surface Extraction In step 1-4 the sought after surface is represented as the zero-level iso-surface of the regularized distance. A polygonized version can be extracted using for example marching cubes [23]. Marching cubes have a tendency to create triangles with bad aspect ratios. For simulation purposes a regular meshing is very important and therefore a mesh optimization scheme is applied. We use a modified version of the technique described in [24] that is enhanced by the use of the implicit description of the surface leading to easy calculations of tangents [17].

RESULTS

The semi-automatic algorithm has been tested on 6 children aged 10 months to 9 years. The result is shown for a single individual, however the remaining results can be obtained by contacting the authors. The result for a nine year old male individual can be seen in Fig. 7. The surface representation was obtained using 12 3D surface scans captured from several directions. When dealing with young children it can not be expected that the head pose is fixed during acquisition, however it is assumed that areas like the pinna and the top of the skull is consistent. The surface representation is therefore build from a single scan in anatomical areas with large interscan variation and from several scans in consistent areas. The obtained results are highly satisfactory since the surface reconstruction looks anatomically correct. The algorithm produces a nice surface in areas with good quality input data and it interpolates in a plausible way in areas with no input samples. The task of choosing the input areas from the 3D surface scans is manual, which results in high-quality and controlled surface representations. Furthermore, the algorithm produces surfaces consisting of near equilateral triangles, as seen in Fig. 7.

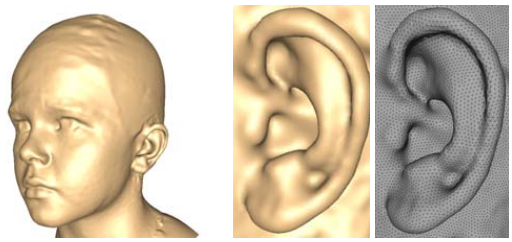


FIGURE 7: Result for a nine year old male individual. Left: Surface reconstruction. Centre: Ear region. Right: Wire frame.

CONCLUSION

A framework for scanning, stitching and meshing complete human heads have been presented. A scan protocol have been developed, in order to assure that the 3D surface scans cover the entire head and ear region and to assure an effective scan session. A semi-automatic method resulted in high-quality and controlled surface representations. The results were highly satisfactory since the surface representations were anatomically correct and the surfaces consisted of near equilateral triangles. Data, scan protocol, algorithm parameters, and software are made available and can be acquired by contacting the corresponding author.

ACKNOWLEDGMENTS

The authors would like to thank the Oticon Foundation for funding this work as well as Thomas Werge and Thomas Hansen for letting us use the 3D surface scanner sponsored by grants to Thomas Werge from The Danish National Advanced Technology foundation and Lundbeck Foundation. Furthermore we would like to thank the staff at Enheden for Psykiatrisk Forskning, Aalborg Psykiatriske Sygehus for hospitality in connection with the scan session.

REFERENCES

[1] J. Blauert, *Spatial Hearing-Revised Edition: The Psychophysics of Human Sound Localization* (MIT press) (1996).

[2] D. Brungart and W. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions", The Journal of the Acoustical Society of America **106**, 1465 (1999).

[3] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements", in *Proc. First IEEE Pacific-Rim Conf. on Multimedia*, 235–238 (Citeseer) (2000).

[4] R. Bentler, C. Pavlovic, *et al.*, "Transfer functions and correction factors used in hearing aid evaluation and research.", Ear and hearing **10**, 58 (1989).

[5] R. Bentler and C. Pavlovic, "Addendum to" transfer functions and correction factors used in hearing aid evaluation and research"", Ear and Hearing **13**, 284–286 (1992).

[6] A. Sontacchi, M. Noisternig, and P. Höldrich, "An objective model of localisation in binaural sound reproduction systems", AES: Proceedings (2012).

[7] S. Jelfs, J. Culling, and M. Lavandier, "Revision and validation of a binaural model for speech intelligibility in noise", Hearing research **275**, 96–104 (2011).

[8] J. Fels, P. Buthmann, and M. Vorlander, "Head-related transfer functions of children", Acta Acustica united with Acustica **90**, 918–927 (2004).

[9] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The cipic hrtf database", in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 99–102 (IEEE) (2001).

[10] P. Guillon, R. Zolfaghari, N. Epain, A. van Schaik, C. Jin, C. Hetherington, J. Thorpe, and A. Tew, "Creating the sydney york morphological and acoustic recordings of ears database", in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, 461–466 (IEEE) (2012).

[11] T. Cootes, C. Taylor, D. Cooper, J. Graham, *et al.*, "Active shape models-their training and application", Computer vision and image understanding **61**, 38–59 (1995).

[12] DTU Compute, "Surface manipulation and transformation toolkit (sumatra)", `http://www2.imm.dtu.dk/~rapa/software.html`.

[13] Visual Computing Lab ISTI - CNR, "Meshlab", `http://meshlab.sourceforge.net/`.

[14] B. Horn, "Closed-form solution of absolute orientation using unit quaternions", JOSA A **4**, 629–642 (1987).

[15] R. Kolluri, J. Shewchuk, and J. O'Brien, "Spectral surface reconstruction from noisy point clouds", in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, 11–21 (ACM) (2004).

[16] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction", in *Proceedings of the fourth Eurographics symposium on Geometry processing*, 61–70 (Eurographics Association) (2006).

[17] R. Paulsen, J. Bærentzen, and R. Larsen, "Markov random field surface reconstruction", Visualization and Computer Graphics, IEEE Transactions on **16**, 636–646 (2010).

[18] V. Caselles, G. Haro, G. Sapiro, and J. Verdera, "On geometric variational models for inpainting surface holes", Computer Vision and Image Understanding **111**, 351–373 (2008).

[19] R. Paulsen and R. Larsen, "Anatomically plausible surface alignment and reconstruction", in *Proceedings of Theory and Practice of Computer Graphics* (2010).

[20] P. J. Besl and N. McKay, "A method of registration of 3D shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence **14**, 239–256 (1992).

[21] M. Jones, J. Bærentzen, and M. Sramek, "3D Distance Fields: A Survey of Techniques and Applications", IEEE Transactions On Visualization and Computer Graphics **12**, 518–599 (2006).

[22] J. Besag, "On the statistical analysis of dirty pictures", Journal of the Royal Statistical Society, Series B **48**, 259–302 (1986).

[23] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm", Proc. SIGGRAPH 163–169 (1987).

[24] M. Botsch and L. Kobbelt, "A remeshing approach to multiresolution modeling", Proc. Symposium on Geometry processing 185–192 (2004).

A.1 Additional material

Complete head scanning

This is a guide on how to scan entire heads using the Canfield Vectra M3 scanner.

Initial front facial shot

Place the person so the preview images looks similar to the ones below (and even with the person being further down). The person should have a neutral expression. If the person does not have very short hair a plastic hairband should be used to remove hair from the forehead.

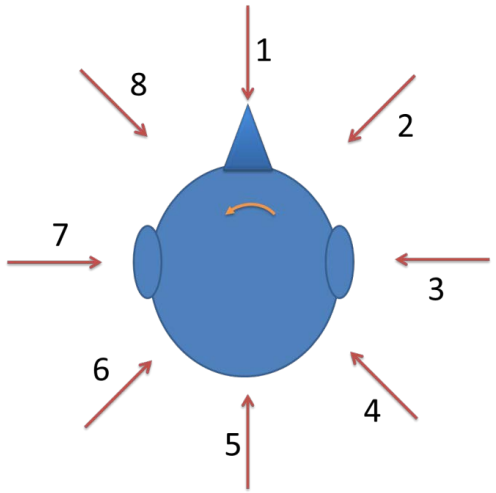
Take two or three shots and keep in mind that one could be used on a personal homepage.

Complete head scan

Carefully cover the subject's hair with a wig cap and place the person on a rotating chair.

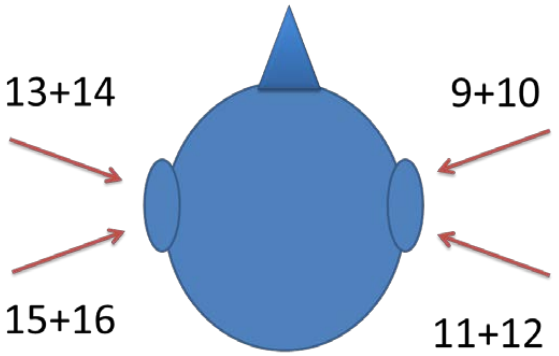
Start by acquiring one or **several** good frontal shots. The subject should look like below on the preview images. This is the reference image so it should be very good:

Acquire 8 scans of the subject while the subject is being as immobile as possible. The scans should be taken from the directions as indicated below.



Ear scans

Take four scans of each ear. Two scans where the head is tilted towards the camera and two where the head is tilted away from the camera. In scan 9, 11, 13, and 15 the head is tilted toward the camera and in scan 10, 12, 14, and 16 the head is tilted away from the camera.



3D full head scanning			
Date			
Place			
Subject name			
Date of birth			
Gender			
Scan library / ID			
Scan #	**Description**	**Cap**	**Done**
	Front facial no cap		
1	Front facial (several)	X	
2	Frontal right side	X	
3	Right ear	X	
4	Back right side	X	
5	Back	X	
6	Back left side	X	
7	Left ear	X	
8	Frontal left side	X	
9	Right ear forward. Tilt towards camera	X	
10	Right ear forward. Tilt away from camera	X	
11	Right ear back. Tilt towards camera	X	
12	Right ear back. Tilt away from camera	X	
13	Left ear forward. Tilt towards camera	X	
14	Left ear forward. Tilt away from camera	X	
15	Left ear back. Tilt towards camera	X	
16	Left ear back. Tilt away from camera	X	

Appendix B

A pipeline of geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions

A pipeline of geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions

Stine Harder[a], Rasmus R. Paulsen[a], Martin Larsen[b], Søren Laugesen[c], Michael Mihocic[d], Piotr Majdak[d]

[a] *Technical University of Denmark, DTU Compute, Department of Applied Mathematics and Computer Science, DK-2800 Lyngby, Denmark*
[b] *Oticon A/S, DK-2765 Smørum, Denmark*
[c] *Eriksholm Research Centre, DK-3070 Snekkersten, Denmark*
[d] *Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria*

Abstract

Individual head-related transfer functions (HRTFs) are essential in applications like fitting hearing-assistive devices (HADs) for providing accurate sound localization performance. Individual HRTFs are usually obtained through cumbersome acoustic measurements. Here, we investigate the use of an individual three-dimensional (3D) head model for acquisition of individual HRTFs. Two aspects were investigated; whether a 3D-printed model can replace measurements on a human listener and whether numerical simulations can replace acoustic measurements. For this purpose, HRTFs were acoustically measured for four human listeners (including repeated measurements) and for a 3D printed head model of one of these listeners. Further, HRTFs were simulated by applying the finite element method to the 3D head model. The monaural spectral features were similar between the acoustic repetitions, somewhat similar between the simulation and measurement of the printed model, but less similar between the model and actual human. The latter differences seem to stem from the different torso geometries. The binaural cues were in agreement among all HRTFs of the same listener, indicating that the 3D model is able to provide localization cues potentially accessible to HAD users. Hence, the pipeline of geometry acquisition, printing, and acoustic measurements or simulations, seems to be a promising step forward towards in-silico design of HADs.

Keywords: Head-related transfer function, acoustical measurements, acoustical simulations, 3D head models, 3D printing

1. Introduction

Human listeners are able to localize sounds in space in terms of assigning direction and distance to the perceived auditory image [1]. This ability is an essential application of spatial hearing, which involves further perceptual effects like the estimation of the apparent source width, spatial unmasking of speech (cocktail party effect), and externalization (out-of-head perception) of sounds. Generally, spatial hearing relies on physical directional acoustic features which are the consequence of acoustic filtering of the sound by the pinna, head, and torso. The filtering can be described by the head-related transfer functions (HRTFs), which represent the transfer functions from sound sources to a sound receiver, usually placed at the entrance of the ear canal, normalized with a free-field response. For far-field sources, HRTFs depend on the direction of sound incidence [2]. For near-field sources, HRTFs additionally depend on the distance between the source and the ear [3].

An HRTF encodes monaural spectral cues, which are used by human listeners to estimate the sound-source position along sagittal planes (top, down, front, back). For the localization of the source along horizontal planes (left, right), interaural cues are used, namely, interaural time and level differences (ITDs, ILDs). In particular, ITD cues in the frequency range below 1.5 kHz are important for the sound localization in the horizontal planes. The interaural cues are encoded in a binaural pair of HRTFs, thus, a binaural set of HRTFs can be used to describe all directional cues required for spatial hearing.

HRTFs can be used for many purposes, e.g., for models of spatial hearing [4, 5, 6, 7], for fitting of hearing aids [8, 9], and for presenting virtual binaural audio signals via headphones in so-called virtual auditory displays [10, 11]. As HRTFs depend on the individual geometry of the listener's head and ear, HRTFs are listener-specific [12]. Listener-specific HRTFs are often acoustically measured by placing small microphones at the entrance of the listener's ear canals. During the measurement, usually done for many spatial positions, the listener must sit still for tens of minutes [13]. Thus, it is not surprising that in some applications, generic HRTFs, i.e., HRTFs of a manikin representing an average of the human population [14], are used. However, in applications like fitting hearing aids to children [15] or providing accurate sound localization performance via headphones [16], it might be important to consider listener-specific HRTFs. As an alternative to demanding acoustic measurements, HRTF can be also numerically calculated from a three-dimensional (3D) representation of human geometry and established methods have shown adequate results for frequencies up to a few kHz [17, 18, 19].

In this study, we evaluated a method for obtaining listener-specific HRTFs with the focus on virtual product design. Virtual product design is an emerging discipline that has the potential of reducing production costs and creating more comfortable and better functioning wearables as for example clothes, helmets, and in our case, hearing-assistive devices (HAD), which are commonly used

to treat hearing impairment. An example of virtual product design is the size-China project that aims at creating a population statistics on human heads for product design [20]. In [21], a parametric model of the entire human body was computed based on 250 full body scans. This model was then used to synthesize plausible body shapes as input to product design. In another example, foot shape was investigated in several studies driven by the large footwear industry and recently, 50 surface scanned feet were used as input to a statistical analysis of shape with the goal to produce optimized shoe lasts [22]. Similarly, in the design of HADs, the position of the microphones is essential, not only for the design of the casings, but also for capturing spatial acoustic cues. Thus, different types of microphones at various places have been proposed, for a recent review, see [23]. The measurement of HRTFs with HADs is even more demanding than the measurement done in the ear canal because of the much larger degree of freedom: A simple re-positioning of the HAD might yield different HRTFs, and many microphone positions have to be considered in the individualization of the HAD. On the other hand, since most of the HADs focus on transmitting frequencies below 8 kHz [24], the evaluation of HRTFs can be limited to the upper frequency 8 kHz.

In this paper, we present yet another step towards full in-silico design of HADs. Our goal was to evaluate whether simulated HRTFs of HADs can replace expensive acoustic measurements in the future. We focused on the issue of re-positioning the HAD's microphones, directly addressing the needs of the hearing-aid industry in the process of product design. The main motivation for that issue was that the re-positioning of the microphone in the simulation does not require additional participation of the listener. HRTF simulations can have a further impact on the future product design: by using statistical shape modeling (as in for example, [21, 22]), HRTFs of future HADs could be obtained for hundreds or even thousands of plausible virtual heads, in contrast to the current practice where one or a few generic manikin heads are used because of the time-consuming measurement procedure.

In the following, we describe the three stages of our investigations. First, we used a high-quality 3D geometrical model of a human listener to numerically simulate corresponding HRTFs. This step was based on the framework for generating listener-specific 3D head-and-ear models from surface scans [25]. Second, we used computer-aided-design modeling and 3D printing to obtain a listener-specific 3D printed head model, from which acoustically measured HRTFs were obtained. Third, we used acoustical measurements to obtain HRTFs of the corresponding actual human listener. This three-step pipeline allowed for a thorough evaluation of the differences among acoustically measured and simulated HRTFs of the model and the listener. The acoustically measured HRTFs for the printed head model made it possible to distinguish between HRTF deviations caused by differences in shape (human versus obtained 3D model) and differences in method (measurements versus simulations).

2. Material and methods

2.1. The surface model

A complete human head model was generated using a framework which combines a number of 3D surface scans captured from different angles surrounding the individual. Listener NH167 from the ARI database (`http://sofaconventions.org`) was scanned. We presented the method in [25] and it consists of both scanning, stitching, and meshing. Here, we briefly review that method.

Sixteen 3D surface scans were acquired using a 3D facial scanner (Canfield scientific Vectra M3) and subsequently aligned with a rigid-body transformation based on a few manually annotated landmark points. The result was a roughly aligned point cloud where many parts of the head were well covered by samples, but several regions lacked well-defined sample points.

Surface reconstruction was applied on the point cloud. It was based on a higher-order mathematical model [26], which is well-suited for human head scans because it 1) is robust to noise, 2) interpolates in areas with missing data in an anatomically plausible way, and 3) accurately aligns the partially overlapping scans. Inspired by [27], the surface reconstruction method consists of several steps that are iterated until convergence. Initially, a single oriented point set with consistent normal directions was created from the input data. Then, a signed distance field was computed based on the oriented point set. In the next step, a Markov random field based regularization was applied to the distance field. In the last step, an implicit variant of the iterative closest point algorithm (ICP) [28] was used to align the input point sets to the implicit zero level. These steps were then repeated in a multi-scale framework. Finally, an iso-surface extraction algorithm was used to extract the resulting surface from the regularized distance field, and the resulting mesh was optimized to obtain a mesh with regular triangles.

The final surface model seems to be anatomically correct and thus satisfactory for further processing, see Figure 1a.

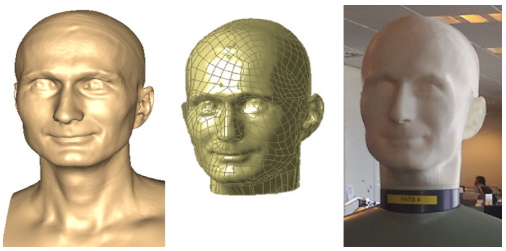


Figure 1: Head-and-ear model. a) Acquired geometry b) Model used for printing c) Printed model mounted on a torso simulator.

2.2. Adapting 3D surface models to simulation and printing

By nature, the 3D surface model acquired from range scanning is a two-dimensional (2D) surface embedded in 3D space. Technically it consists of vertices and triangles, which also applies to the results of the surface merging and stitching. Several numerical simulation packages exist that can use triangulated surfaces as input but most programs

work better with surfaces represented as non-uniform rational B-splines (NURBS) [29]. Thus, the surface model was converted to NURBS (Geomagic Studio, www.geomagic.com), see Figure 1b.

The head model represented in NURBS was further processed using a computer-aided design software package (Creo Parametric 2.0). Because the surface model has no thickness per se, an artificial thickness was added to the surface. This is typically done by creating an inward offset surface of several millimeters. However, with the complicated geometry of the human head and ears this approach did not give satisfactory results. Instead, a balloon-like object was created inside the head and manually reshaped to turn the head into a hollow shell. The overall minimum thickness was 5 millimeter and certain parts like the nose remained solid.

In order to mount a torso simulator (Type 4128, Brüel & Kjær) after the printing, a flange was added to the model at the bottom of the head. Further, in order to be able to manufacture the head and ears separately using different materials, boxes around the ears were created. After the printing, the ears were attached to box-shaped elements that were inserted in the corresponding box-shaped recesses in the head.

2.3. 3D print specifications for the head and ears

With the rapid development in layered manufacturing (in the following 3D printing) new possibilities are opening up. Previously, creating an accurate replica of a human head would have required massive amounts of work. As demonstrated in this article it is still not a trivial task, but with the development of software tools and the decreasing costs of 3D printing, generating flexible human replicas for product testing is becoming increasingly accessible.

In the current work, we have not focused on creating a full head with material properties that exactly match a living human. It would be very complicated and it is believed that the impact on the acoustical measurements is of less significance than other factors in the measurement setup. The head was therefore printed in hard plastic except the ears which were printed in a soft material. The reason for choosing a soft material for the ears was mainly the need of placing hearing aids behind the ears in a way that is as close as possible to the placement of hearing aids on human ears.

The ears were printed on a polyjet Objet500 Connex 3D printer (www.stratasys.com) using two materials and a layer thickness of 0.03 mm. The core was printed in a hard, acrylic material (Verowhite - FullCure830) and the outer part was printed in a soft material (TangoPlus - FullCure930). The hard part of the head was printed on an SLS printer EOS P395 (www.eos.info) in hard acrylic (PA2200) using 0.12 mm layers. Further material properties can be located at (www.damvig.dk). A set of ears were produced where wax residues were removed using a bath of base. However, it seemed that this made the material brittle and further experiments were done without cleaning with a base.

2.4. HRTF simulation

The simulation was set up in ANSYS 15.0. The head geometry from Section 2.2 was mounted on a virtual model of the Brüel & Kjær torso simulator for the simulations. The head geometry was meshed with first-order acoustic elements (ANSYS type 30). The mesh was refined at the ears to an average edge length of 2 mm, while the rest of the model used a length of 4 mm. The re-meshing method was 'Hex dominant', which produced a mesh with less nodes compared to a tetrahedral mesh. The final mesh, which represented the model of the head and torso, contained 1.9 million nodes.

The final model was placed inside a box with dimensions 420 mm×700 mm×250 mm, which in the simulation was set to be filled with air. The model was subtracted from the box, leaving the air surrounding the model. A 40 mm perfectly matched layer (PML) was added to the outside of the inner air-box, see Figure 2. The purpose of the PML was to absorb radiated sound from the inner air part. The default setting for attenuation in the PML region was used (0.001 or -60 dB), yielding 120 dB total attenuation of the sound field reflected from the PML.

In the acoustic measurement setup (see Section 2.5), loudspeakers surrounding the head generated the sound, and the microphones placed in the ear canal captured the sound pressure. In the simulation, the reciprocal approach was used [30, 31]: the sound source was placed at the microphone position, and the sound pressure level was calculated on a sphere surrounding the head at 1.2 m distance. The reciprocal approach yields equivalent results to the direct approach, but the simulation can be performed for all sound-source directions in a single simulation step [32].

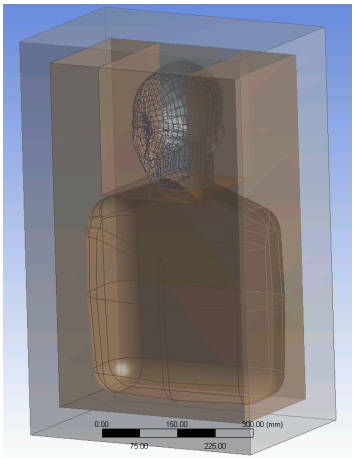


Figure 2: Geometry layers used for HRTF simulation. From inside to out: empty space shaped as the head and torso, inner air box, and perfectly matched layer.

The simulation was performed for a total of 54 linearly spaced frequencies from 187.5 to 10 125 Hz and for the two microphones positioned in the ear canals. The resulting far field was evaluated in vertical and horizontal angle steps of 2.5 degrees at a distance of 1.2 m, however, only the 1550 source positions, that were acquired for the measurements (see Section 2.5), were used for this paper.

Furthermore, a free-field transfer function was simulated and used as a reference for the other simulations, resulting in a set of simulated HRTFs comparable to the measured HRTFs described in Section 2.5. In this way

both measured and simulated HRTFs were ratios of sound pressures at a microphone on the head and at a microphone in the center of the head (head absent).

2.5. HRTF acoustic measurement

2.5.1. Human listeners

HRTFs of human listeners were acoustically measured in a semi-anechoic chamber. Twenty-two loudspeakers (custom-made boxes with VIFA 10 BGS as drivers; the variation in the frequency response was ±4 dB in the range from 0.2 to 16 kHz) were mounted on an arc at fixed elevations from −30 ° to 80 °. They were driven by amplifiers adapted from Edirol MA-5D active loudspeaker systems. The loudspeakers and the arc were covered with acoustic damping material to reduce the intensity of reflections. The total harmonic distortion of the loudspeaker-amplifier systems was on average 0.19 % (at 63-dB SPL and 1 kHz). The human listener was seated on a chair in the center of the arc and was wearing in-ear-microphones (Sennheiser KE-4-211-2). HRTFs were measured at the entrance of the blocked ear canal. The microphones were connected via amplifiers (RDL FP-MP1) to the digital audio interface. A 1728.8-ms exponential frequency sweep beginning at 50 Hz and ending at 20 kHz was used to measure each HRTF. Then a system identification procedure was used to obtain raw HRTFs [13].

The HRTFs were measured for one azimuth and several elevations at once (see below) by playing the sweeps and recording the signals at the microphones. Then the listener was rotated by 2.5 ° to measure HRTFs for the next azimuth. In the horizontal interaural plane, the HRTFs were measured with 2.5 ° spacing within the azimuth range of ±45 ° and with 5 ° spacing outside this range. The positions of the HRTFs were distributed with an approximately constant spherical angle, which means that the number of measured HRTFs in a given horizontal plane decreased with increasing elevation. For example, at the elevation of 80 °, only 18 HRTFs were measured. In total, 1550 HRTFs were measured for each listener. To decrease the total time required to measure the HRTFs, the multiple exponential sweep method (MESM) was applied [13]. This method allows for a subsequent sweep to be played before the end of a previous sweep, but still reconstructs HRTFs without artifacts. The MESM uses two mechanisms, interleaving and overlapping and both depend on the acoustic measurement conditions (for more details see [13]). Our facilities allowed the interleaving of three sweeps and overlapping of eight groups of the interleaved sweeps. During the HRTF measurement, the position and orientation of the listener's head were captured via an electromagnetic tracker (Flock of Birds, Ascension) in real time. The tracking sensor was mounted on the top of the listener's head. The tracking device was capable of measuring all 6 degrees of freedom (x, y, z, azimuth, elevation, and roll) at a rate of 51.5 measurements/sec. The tracking accuracy was 1.7 mm for positions and 0.5 ° for orientation. If the head was outside the valid range, the measurements for that particular azimuth were repeated once the listener was back in the range for 500 ms. The

valid ranges were set to ± 2.5 cm for the position, ± 2.5 ° for the azimuth, and ± 5 ° for the elevation and roll. On average, measurements for three azimuths were repeated per listener and the complete measurement procedure lasted for approximately 20 minutes.

For each raw HRTF, head-related impulse responses (HRIRs), i.e., the inverse Fourier transform of HRTFs, was calculated, yielding raw HRIRs. The raw HRTFs were affected by loudspeaker, room, and microphone used for the measurements. These effects can be described by the equipment transfer functions (ETFs), which were captured by placing the in-ear microphones in the center of the arc and measuring the room impulse response for all loudspeakers. ETFs were measured each time before the HRTF measurement of an object, thus five times in total.

The impulse responses of the ETFs showed a strong reflection of at least 20 dB below the level of the direct sound, delayed by at least 6.9 ms. This reflection can be attributed to the floor reflection as an effect of using a semi-anechoic room for the measurements. Since HRIRs do not contribute much beyond first 5 ms [33], such reflections were removed from both the raw HRIRs and the impulse responses of ETFs by windowing in the time domain.

The remaining loudspeaker and microphone responses were compensated by dividing an HRTF for a given microphone, m, and loudspeaker position, (θ, ϕ), with its corresponding ETF. The set of compensated HRTFs for all directions and for both microphones, will from now on be referred to as the *human* HRTFs.

HRTFs of four listeners (NH166, NH167, NH168, and NH170 from the ARI database, `http://sofaconventions.org`) were measured. In addition, the measurement was repeated for NH167. Thus, five sets of human HRTFs were available for the analysis.

2.5.2. Printed model

HRTFs of the printed head fixed on a torso simulator (Type 4128, Brüel & Kjær) were measured using the same methodology as for the human listeners with one exception: The torso was fixed with tape in order to prevent any movements during the measurements. Also, all other aspects of post processing were identical to those of the post processing done for the human listeners, resulting in *printed* HRTFs.

2.6. HRTF evaluation

2.6.1. Monaural spectral features

The monaural spectral features were analyzed by visual comparison of the amplitude spectra of the corresponding HRTFs. The spectra were logarithmically smoothed with a 1/3-octave wide rectangular window to obtain patterns which can be compared more easily. Similar comparisons were performed in studies like [15, 34, 18, 19].

For the evaluation of the ITD cues, the ITD was evaluated as a function of the azimuth angle in the horizontal plane by using a similar approach to that used in [35]. For each azimuth ϕ, the ITD was derived from the interaural phase difference of the corresponding HRTFs. In particular, given a pair of HRTFs, $H_{l,\phi}$ and $H_{r,\phi}$ for a given source

position in the frequency domain, the interaural phase difference, $\Phi_\phi(f)$, was

$$\Phi_\phi(f) = \text{unwrap}\left(\frac{H_{l,\phi}}{H_{r,\phi}}\right), \qquad (1)$$

where unwrap(.) denotes the "unwrap-phase" operator along the frequency axis. Then, the frequency-dependent ITD, $\tau_\phi(f)$, is

$$\tau_\phi(f) = \frac{-\phi_\phi(f)}{2\pi f}. \qquad (2)$$

The estimated ITD, τ_ϕ, was then the $\tau_\phi(f)$ averaged in the frequency range from 164 to 890 Hz for the human and printed HRTFs, and from 188 to 938 Hz for the simulated HRTFs, respectively.

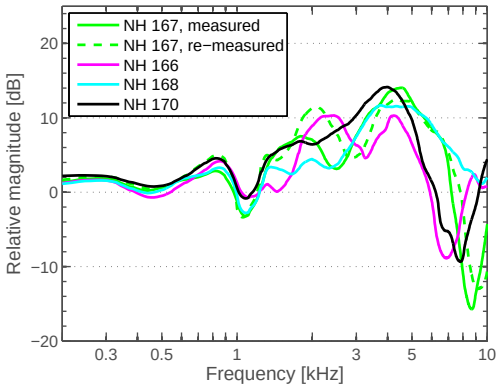


Figure 3: Right-ear HRTF amplitudes for the frontal position in the horizontal plane (elevation and azimuth angles of 0 degrees).

3. Results

3.1. Monaural spectral features

Figure 3 shows the human HRTFs for the frontal position in the horizontal plane (both azimuth and elevation of zero) for the right ear. Even though the general shapes are similar, large variations can be observed among the listeners. The repetition of the HRTF measurement yielded similar HRTFs for frequencies below 1.7 kHz but large variations can be observed for higher frequencies.

Figure 4 shows the human HRTFs of NH167, the repeated human HRTFs, printed HRTFs, and the simulated HRTFs for the right ear. Five positions in the horizontal plane are shown with azimuth angles of 90, 45, 0, -45, and -90 degrees. As expected, the shapes of the HRTFs change with the azimuth angle for all types of HRTFs. For the ipsilateral directions (negative azimuth angles), the human, printed, and simulated HRTFs show a good similarity, especially in the frequency range up to 4 kHz. A generally larger difference between the HRTFs can be observed for the contralateral directions (positive azimuth angles). Generally, differences are small at lower frequencies and increase with frequency.

For the frontal direction (azimuth of zero) an interesting observation can be made. The two human HRTFs, and the printed and simulated HRTFs are in a good agreement with each other. While the human HRTFs are more different from both the printed and simulated HRTFs. This observation, namely small differences between the human HRTFs, small differences between the printed and simulated HRTFs, but, larger differences between human and printed and/or simulated HRTFs applies more generally in other az-

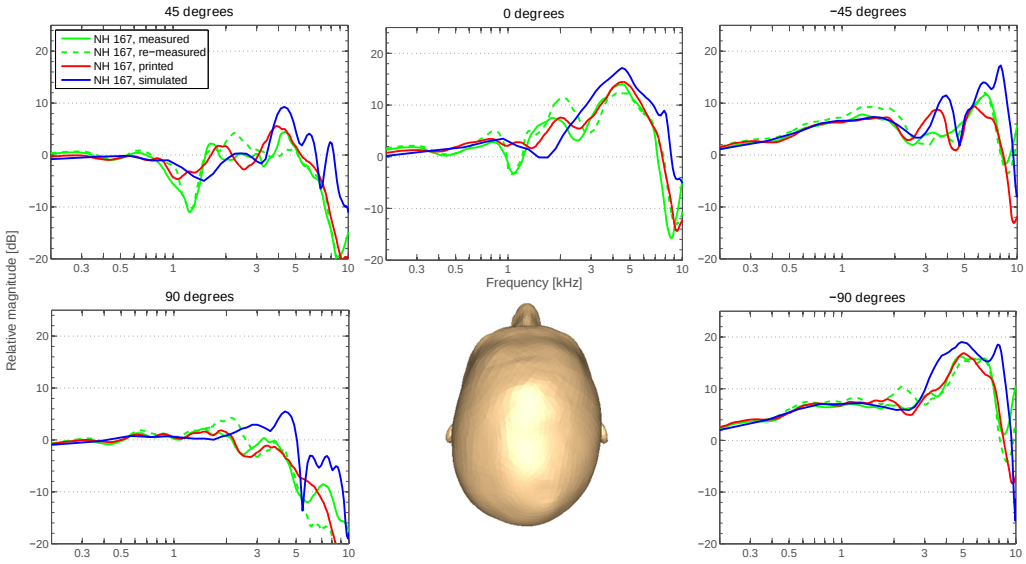


Figure 4: Right-ear HRTF magnitude spectra in the horizontal plane (elevation angle of 0 degrees) and the azimuth angle of 90, 45, 0, -45 and -90 degrees.

imuthal angles as well. For example, for the azimuth of 45 and -45 degrees, the peaks in the printed HRTFs were captured by the simulation (even though they appear to be shifted to a higher frequency) but they are much less pronounced in the human HRTFs. Also, a large difference between the human and printed HRTFs can be observed at approximately 1.1 kHz for azimuth angles of 45 and 0 degrees, where a notch is present in the human but missing in the printed HRTFs. The frequency of 1.1 kHz corresponds to a wavelength of 30 cm and the notch can therefore be attributed to differences between the torso of the actual human and the torso used for the measurements with the printed head. Recall that the printed head was fixed on a torso simulator, which had a different shape and probably different acoustic properties in terms of surface impedance. The reflections caused by the torso manifest themselves as comb-filter effects in the amplitude spectra. With the different torsos, differences in the comb-filter effect are apparent and also the peak at 3.4 kHz seems to be part of that effect.

In order to address that issue, the amplitudes of HRTFs were plotted as a function of the elevation angle in the median plane, i.e., for all azimuth angles of zero degrees, see Figure 5. Human, printed, and simulated HRTFs are shown, the color encodes the rela-

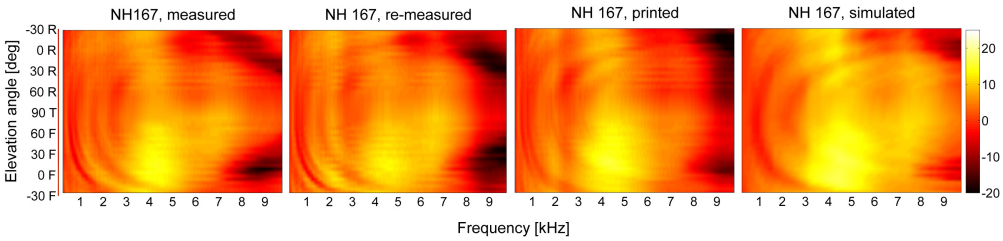


Figure 5: HRTF magnitude spectra in the median plane (as function of elevation angle). Color encodes the relative magnitude in dB. F, T, R describe directions in front, top, and rear of the listener.

tive magnitude (in dB). The spectra at elevation of zero degrees correspond to the spectra shown in the panel for azimuth angle of zero in Figure 3.

In all four panels of Figure 5, spectral modulations can be observed, beginning at frequencies around 2 kHz and polar angles of $-30\,^{\circ}$, and then decreasing in frequencies with increasing polar angle up to $0\,^{\circ}$, with an opposite pattern in the rear hemifield. There is, however, a striking difference in these patterns between the human and the printed and/or simulated HRTFs: the modulations in the human HRTFs are more spectrally compressed, with more pronounced modulation depth.

Besides these differences in the patterns, the spectral features appear to be similar across all four HRTF types.

3.2. ITD cues

Figure 6 shows ITDs in the horizontal plane calculated from human, printed, and simulated HRTFs. For NH167, the ITD showed similar patterns when compared across the two acoustic measurements, with a maximum ITD of 790 μs. These ITD patterns also seem to be similar when compared to those obtained from the printed and simulated HRTFs. The ITDs obtained for the remaining three listeners exhibit clear differences with maximum ITDs of 760 μs, 810 μs and 830 μs, respectively. These differences can be attributed to different head sizes of the test subjects.

4. Discussion

The goal of the present study was to investigate the use of an individual head model in the acquisition of individual HRTFs. One aspect was to investigate whether a 3D printed version of a listener's head can replace the human listener in measurements. Another aspect was to evaluate the quality of simulated HRTFs when using an individual head model.

Generally, there was a variability in amplitude spectra of HRTFs across the listeners, particularly for higher frequencies. However, that variability was comparable to that found for repeated measurements on the same

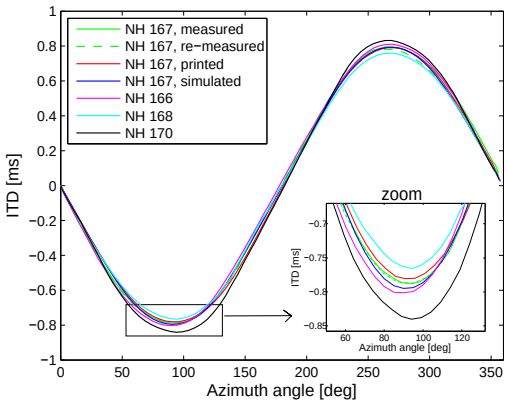


Figure 6: ITDs in the horizontal plane (as function of azimuth angle).The right panel shows a zoom around the most lateral directions shown in the left panel.

listener. The observed variability can be attributed to small changes in the measurement setup, namely, 1) microphone position (recall that microphone was re-positioned between the two measurements), and 2) position and orientation of the listener (recall the criteria for accepting a measurement, see Section 2.5.1). The acoustic consequences of these geometrical changes naturally become greater at high frequencies where the wavelength of sound is shorter and thus in the range of the magnitude of the geometrical changes. The variability captured by the repeated measurement sets a reference for further evaluation of the differences among the human, printed, and simulated HRTFs of that listener. Interestingly, similar variability between two HRTF measurements in the same listeners was found in([36], Fig. 3 conditions binOwn and binOwnB). Despite the differences in the HRTFs, sound-localization performance obtained with these two conditions in localization experiments was similar. This indicates that also our repeated measurements are similar in terms of providing similar directional cues.

A variability was also found for the ITDs, the most salient cue for localizing sound sources in the horizontal plane. However, the ITD variability showed a different pattern than the monaural amplitude spectra: very similar ITDs were found for all HRTFs of the same listener NH167 (smaller than 30 μs, which is below the ITD discrimination threshold found at that direction, [37]), but different ITDs were found across the listeners. Since the shape of the ITD as a function of azimuth theoretically depends on the shape and size of the listener's head, it was not surprising that we also found across-listener differences.

The question of whether a 3D printed version of a listener's head can replace repeated measurements of the human listener was addressed in the comparison between the human and printed HRTFs. The comparison of the amplitude spectra revealed some differences. Some of these differences, when compared to those obtained by the repeated measurement on the actual human, have been attributed to small differences in microphone position and head orientation. Given the limited frequency range of the HADs, these differences can be considered as minor. Other differences have been attributed to differences in the torso used for the measurements with the actual human and the printed model.

These differences were reflected as differences in spectral amplitude modulations as a function of the elevation angle, and can thus be interpreted as modified encoding of the directional position of the sound source. Nevertheless, such modifications will most probably not affect the sound-localization ability because ITDs did not show such modifications, and because torso reflections represent only a weak cue for sound localization [38]. However, the different spectral modulations might have some effect on non-directional quality aspects of hearing like timbre. Thus, in order to match the timbre as well as directional cues, an individualization of the torso geometry might be required.

The aspect of the quality of simulated HRTFs was addressed in the comparison between the printed and simulated HRTFs. Generally, printed and simulated HRTFs appear to be similar (in terms of both ITDs and amplitude), indicating a good match between the acoustical measurement of the printed copy of a listener and the simulations. Note that the shape of the torso used in the measurement of the printed head was exactly translated to the simulations. However, some local deviations between the HRTFs can be observed, especially at frequencies above 3 kHz. While the head and ear model used for printing was exactly translated to the simulation, some errors might have been introduced by the printing process. The layer thickness and thereby printing accuracy was, as stated in Section 2.3, 0.03 mm and 0.12 mm for the ears and head, respectively. Thus, the printing errors rather unlikely contributed to the noticeable HRTF differences between measurement and simulation. The microphone position, however, might have been more critical. Even though photo documentation of the microphone position was captured for the measurements and used to guide the positioning of the microphone in the simulation, the position of the microphone might have been translated with a deviation in the range of millimeters. Further differences between the HRTFs can also arise due to differences between measuring and simulating HRTFs. For example, the measurements were performed in a semi-anechoic chamber with electro-acoustic equipment involved, whereas the simulations had a PML that absorbed all radiated sound. Further, the printed head and torso simulator was attached to a chair during the measurements, whereas they were perfectly (virtually) placed in the simulations.

5. Conclusions

Similar monaural cues were shown between the repeated human measurements for subject NH167. Somewhat similar monaural cues were observed between the 3D model measurements and simulations. However, less similarities, especially in the 1 kHz range, were observed between human and model measurements and simulations. The differences were mainly attributed to differences in torso geometry. The ITDs revealed a good agreement among the three different kinds of HRTFs for subject NH167, whereas meaningful differences were observed among listeners. The very similar ITDs observed for subject NH167 indicate that a printed head model

can be used for obtaining listener specific spatial binaural cues.

6. Bibliography

[1] J. Strutt, Our perception of the direction of a source of sound, Nature 14 (1876) 32–33.

[2] J. Blauert, Spatial Hearing-Revised Edition: The Psychophysics of Human Sound Localization, MIT press, 1996.

[3] D. Brungart, W. Rabinowitz, Auditory localization of nearby sources. head-related transfer functions, The Journal of the Acoustical Society of America 106 (1999) 1465.

[4] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, Z. Tang, Approximating the head-related transfer function using simple geometric models of the head and torso, The Journal of the Acoustical Society of America 112 (5) (2002) 2053–2064.

[5] R. Baumgartner, P. Majdak, B. Laback, Modeling sound-source localization in sagittal planes for human listeners (2014).

[6] W. Gaik, Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling, The Journal of the Acoustical Society of America 94 (1) (1993) 98–110.

[7] S. Jelfs, J. Culling, M. Lavandier, Revision and validation of a binaural model for speech intelligibility in noise, Hearing research 275 (1) (2011) 96–104.

[8] R. Bentler, C. Pavlovic, et al., Transfer functions and correction factors used in hearing aid evaluation and research., Ear and hearing 10 (1) (1989) 58.

[9] R. Bentler, C. Pavlovic, Addendum to" transfer functions and correction factors used in hearing aid evaluation and research", Ear and Hearing 13 (4) (1992) 284–286.

[10] W. M. Hartmann, A. Wittenberg, On the externalization of sound images, The Journal of the Acoustical Society of America 99 (6) (1996) 3678–3688.

[11] F. L. Wightman, D. J. Kistler, Headphone simulation of free-field listening. i: Stimulus synthesis, The Journal of the Acoustical Society of America 85 (2) (1989) 858–867.

[12] F. L. Wightman, D. J. Kistler, Factors affecting the relative salience of sound localization cues, Binaural and spatial hearing in real and virtual environments 1 (1997) 1–23.

[13] P. Majdak, P. Balazs, B. Laback, Multiple exponential sweep method for fast measurement of head-related transfer functions, Journal of the Audio Engineering Society 55 (7/8) (2007) 623–637.

[14] W. Gardner, Martin. kd hrtf measurements of a kemar, Journal of the Acoustical Society of America 97 (6) (1995) 3907–3908.

[15] J. Fels, P. Buthmann, M. Vorlander, Head-related transfer functions of children, Acta Acustica united with Acustica 90 (5) (2004) 918–927.

[16] J. C. Middlebrooks, Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency, The Journal of the Acoustical Society of America 106 (3) (1999) 1493–1510.

[17] Y. Kahana, P. A. Nelson, Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models, Journal of sound and vibration 300 (3) (2007) 552–579.

[18] B. F. Katz, Boundary element method calculation of individual head-related transfer function. i. rigid model calculation, The Journal of the Acoustical Society of America 110 (5) (2001) 2440–2448.

[19] W. Kreuzer, P. Majdak, Z. Chen, Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range, The Journal of the Acoustical Society of America 126 (3) (2009) 1280–1290.

[20] Y. Luximon, R. Ball, L. Justice, The 3d chinese head and face modeling, Computer-Aided Design 44 (1) (2012) 40–47.

[21] S.-Y. Baek, K. Lee, Parametric human body shape modeling framework for human-centered product design, Computer-Aided Design 44 (1) (2012) 56–67.

[22] A. S. Rodrigo, R. S. Goonetilleke, C. P. Witana, Model based foot shape classification using 2d foot outlines, Computer-Aided Design 44 (1) (2012) 48–55.

[23] R. W. McCreery, R. A. Venediktov, J. J. Coleman, H. M. Leech, An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss, American journal of audiology 21 (2) (2012) 295–312.

[24] M. Valente, Hearing aids: Standards, options, and limitations, Thieme, 2002.

[25] S. Harder, R. R. Paulsen, M. Larsen, et al., A three dimensional children head database for acoustical research and development, in: Proceedings of Meetings on Acoustics, Vol. 19, Acoustical Society of America, 2013, p. 050013.

[26] R. R. Paulsen, J. A. Bærentzen, R. Larsen, Markov random field surface reconstruction, Visualization and Computer Graphics, IEEE Transactions on 16 (4) (2010) 636–646.

[27] R. Paulsen, R. Larsen, Anatomically plausible surface alignment and reconstruction, in: Proceedings of Theory and Practice of Computer Graphics, 2010.

[28] P. J. Besl, N. D. McKay, Method for registration of 3-d shapes, in: Robotics-DL

tentative, International Society for Optics and Photonics, 1992, pp. 586–606.

[29] J. D. Foley, A. Van Dam, S. K. Feiner, J. F. Hughes, R. L. Phillips, Introduction to computer graphics, Vol. 55, Addison-Wesley Reading, 1994.

[30] P. M. Morse, Theoretical acoustics, Princeton University Press, 1986.

[31] D. N. Zotkin, R. Duraiswami, E. Grassi, N. A. Gumerov, Fast head-related transfer function measurement via reciprocity, The Journal of the Acoustical Society of America 120 (4) (2006) 2202–2215.

[32] F. Jacobsen, P. M. Juhl, Fundamentals of General Linear Acoustics, John Wiley & Sons, 2013.

[33] M. A. Senova, K. I. McAnally, R. L. Martin, Localization of virtual sound as a function of head-related impulse response duration, Journal of the Audio Engineering Society 50 (1/2) (2002) 57–66.

[34] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, D. N. Zotkin, Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation, The Journal of the Acoustical Society of America 127 (1) (2010) 370–386.

[35] M. Rébillat, V. Benichoux, M. Otani, R. Keriven, R. Brette, Estimation of the low-frequency components of the head-related transfer functions of animals from photographs, The Journal of the Acoustical Society of America 135 (5) (2014) 2534–2544.

[36] P. Majdak, B. Masiero, J. Fels, Sound localization in individualized and non-individualized crosstalk cancellation systems, The Journal of the Acoustical Society of America 133 (4) (2013) 2055–2068.

[37] R. Domnitz, The interaural time jnd as a simultaneous function of interaural time and interaural amplitude, The Journal of the Acoustical Society of America 53 (6) (1973) 1549–1552.

[38] V. Algazi, R. Duda, D. Thompson, C. Avendano, The cipic hrtf database, in: Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, IEEE, 2001, pp. 99–102.

Appendix C

Reliability in measuring head related transfer functions of hearing aids

Acta Acustica united with Acustica
Vol. 98 (2012)

Reliability in measuring head related transfer functions of hearing aids

S. Harder[1)], R. R. Paulsen[1)], M. Larsen[2)], S. Laugesen[3)], M. Mihocic[4)], P. Majdak[4)]
[1)] Technical University of Denmark, DTU Compute, Department of Applied Mathematics and Computer Science, DK- 2800 Lyngby, Denmark. sthar@dtu.dk
[2)] Oticon A/S, DK-2765 Smørum, Denmark.
[3)] Eriksholm Research Centre, DK-3070 Snekkersten, Denmark.
[4)] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria.

Summary
The present paper investigates the reliability of HRTFs measurements for hearing aids. We use a 3D printed head model for HRTF measurements with three different conditions: simple repetitions, repositioning of hearing aids, and repositioning of printed head as well as repositioning of hearing aids. The simple repetition shows very reliable measurements, with standard deviations in the range 0.01 dB. A repositioning of the hearing aids shows less reliable measurements, with standard deviations up to 0.6 dB. The repositioning of both printed head and hearing aids does not show a substantial increase in the variability compared to only repositioning the hearing aid. We also measure HRTFs in different sessions for both the printed head and a human test subject. The results show that HRTFs measured in different sessions have a larger variability than those measured within a session, attributed to less controlled measurement conditions. Furthermore, measurements on a human subject contain more variability than measurements on the printed head, which is likely to be caused by small unavoidable head or body movements.

PACS no. 43 ...

1. Introduction

Head related transfer functions (HRTFs) are a numerical representation of spatial acoustic cues. They are important in applications like modeling spatial hearing [1, 2]. Since HRTFs depend on the listener's head and ear, HRTFs are listener-specific [3]. Individual HRTFs are especially important for applications like fitting hearing aids to children [4] or providing accurate sound localization performance via headphones [5]. In [6] we investigated the use of a 3D printed head model for obtaining listener-specific HRTFs. The goal was to investigate how close to the true HRTF (measured on a human test subject) one could get, by measuring and simulating HRTFs on a 3D printed head and a virtual 3D head model, respectively.

The reliability of HRTF measurements was investigated by Møller et al. [7] and Wightman et al. [8]. Møller et al. showed good agreement between HRTFs measured under the same conditions at different days and Wightman et al. showed the variability in HRTFs caused by head movements and microphone placement. In both cases, HRTFs were obtained for a human test subject with a microphone placed in the ear canal. The reliability of HRTFs measured with hearing aids is less well explored.

Hence, in this study, HRTF measurement reliability with microphones placed in hearing aids was investigated. Further, we used a 3D printed head model as a supplement to the human test subject, exploiting the benefit of excluding head movements usually present when measuring human listeners. A listener-specific head model might replace time-consuming measurements on a human test subject, and it is therefore interesting to know the reliability of such measurements, and to compare it with the reliability of HRTF measurements on human test subjects

Received 27 October 2012,
accepted 6 December 2012.

2. Material and methods

2.1. Generation of 3D printed head

A listener-specific 3D head and ear model was generated using the framework in [9]. The 3D model was based on a number of 3D surface scans of the human listener NH167 from the ARI database (`http://sofaconventions.org`). The scans were obtained

using a Canfield Scientific Vectra M3 surface scanner and were captured from different angles surrounding the head. The surface scans were then aligned and stitched together using a higher-order mathematical model. The complete 3D head and ear model was post-processed and printed using the method described in [6].

2.2. Acoustical measurements

HRTFs were measured in a semi-anechoic chamber. Twenty-two loudspeakers were mounted on an arc at fixed elevations from −30° to 80°, (for more details on equipment, see [10]). The printed head was fixed on a torso simulator (Type 4128, Brüel & Kjær) and placed in a chair in the center of the arc, such that the HRTFs were measured for an azimuth angle of 0°. The printed head or human subject was wearing hearing-aid shells with microphones on both ears. Two types of hearing-aid shells were used: behind-the-ear (BTE) and completely-in-canal (CIC, for more details on the hearing-aid microphones, see [6]). The microphones were connected via amplifiers to the digital audio interface. A 1728.8-ms exponential frequency sweep beginning at 50 Hz and ending at 20 kHz was used to measure each HRTF. Note that in the analysis, we were interested in the frequency range used by hearing aids, i.e., up to 10 kHz only.

The HRTFs were affected by the loudspeaker, room, and microphone used for the measurements. These effects can be described by the equipment transfer functions (ETFs). ETFs were captured with a reference measurement in which the microphones were placed in the center of the arc and the room impulse response was measured for all loudspeakers. Room reflections were initially removed from both the HRTFs and the ETFs by windowing in the time domain, as it was done in [6].

The remaining loudspeaker and microphone responses were compensated in the frequency domain. For a given microphone k, a compensated HRTF $C_k(\theta, \phi)$ at elevation θ, and azimuth ϕ, was calculated as:

$$C_k(\theta, \phi) = \frac{H_k(\theta, \phi)}{E_F(\theta)}, \qquad (1)$$

where $H_k(\theta)$ is the windowed raw HRTF for microphone k and $E_F(\theta)$ is the windowed ETF for the front microphone in the BTE hearing aid.

In this way the HRTFs have been compensated for room and loudspeaker responses, however they are not compensated for the individual microphone response. The small phase differences that exist between the BTE microphones are highly important for hearing-aid directionality [11]. We realized that the best way to retain the phase difference during the post processing was to use the ETF for one microphone in a set of BTE microphones, instead of compensating each microphone using their respective EFT. We hereby assume that the two matched microphones in a BTE hearing aid have equivalent microphone responses.

2.3. Conditions

The measurements were repeated under various conditions. In the condition *HAD*, hearing aids were detached from the head and re-attached before the next measurement. In the conditions *Tripod1* and *Tripod2$_s$*, the tripod carrying the torso, head, and the hearing aids was removed from the arc center and placed back. All conditions except Tripod2$_s$ were measured five times. Notice that Tripod2$_s$ has a subscript s to indicate a single measurement. Furthermore, hearing aids were detached from the head and re-attached. Tripod1 and Tripod2$_s$ were performed on two different days. In the condition *Control*, the measurements were repeated without any intervention.

As a further control condition, single measurements on the actual human listener were performed on two different days yielding *Human1$_s$* and *Human2$_s$*, respectively. Note that the human listener was the model for the 3D printed head, thus, the listener and the printed head shared the same anatomical details.

All conditions were measured for both CIC and BTE hearing aids with the exception of the condition HAD, which was not measured for the CIC hearing aid because of difficulties with repositioning the hearing aid. The ear canal of the printed head was not as flexible as an ear canal on a human, and insertion and removal of the CIC was difficult to control. Furthermore, for the CIC microphones in condition Tripod1, only the head and torso and not the hearing aid were repositioned.

3. Results and discussion

3.1. Within-session repetitions

Figure 1 shows spectral amplitudes of five HRTF measurements for each of the three conditions: Control, HAD, and Tripod1. The HRTFs are shown for the left-frontal (LF) microphone of the BTE hearing aid and for the exemplary direction of elevation angle of 15° (other elevation angles showed similar results). In the figure, the three conditions are offset by 5 dB for better readability. The top panel shows the absolute amplitudes and the bottom panel shows the amplitudes relative to the average. It appears that from Control over HAD to Tripod1, the measurement repetition had more and more effect. This is in agreement with the measurement conditions, because from Control over HAD to Tripod1, more and more parts have been re-positioned between the measurements. Nevertheless, the variability is below 1 dB in most cases.

Figure 2 which shows HRTF magnitude spectra based on a single measurement for conditions Control, HAD, and Tripod1 measurements. The HRTFs

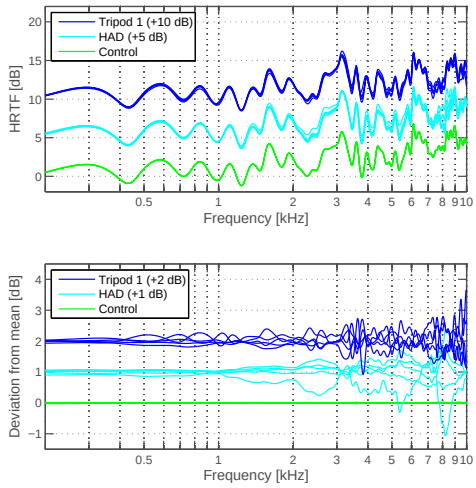


Figure 1. HRTF magnitude spectra for elevation of 15° and conditions Control, HAD, and Tripod1, each repeated five times. Top panel: Absolute magnitudes with an offset of 5 dB between conditions. Bottom panel: Deviation from the average with an offset of 1 dB between conditions for visualization purposes.

seem to be similar in the three conditions for all elevation angles, supporting that the findings for the single elevation shown in detail in Figure 1 are general. In order to quantitatively analyze the variability for all tested directions and hearing aids, spectral standard deviations (SDs) were calculated across the five measurements for each frequency and condition. The average across frequency of these SDs calculated between 0.2 and 10 kHz is shown in Table I. A very small variability is observed in the condition Control, with a standard deviation of 0.01 dB. The obtained HRTFs were practically identical, leading to deviation

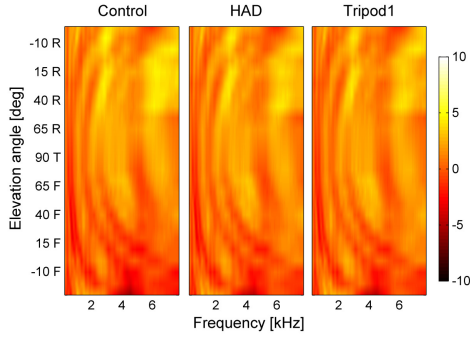


Figure 2. HRTF magnitude spectra in the median plane (as function of elevation angle). Color encodes the relative magnitude in dB. F, T, R indicate the front, top, and rear directions, respectively.

	BTE				CIC	
	LF	LR	RF	RR	L	R
Control	0.01	0.01	0.01	0.01	0.03	0.03
HAD	0.46	0.60	0.36	0.53	-	-
Tripod1	0.34	0.39	0.41	0.72	0.56	0.58

Table I. Standard deviations (in dB) calculated for the five repetitions for each elevation angle and averaged over the frequency range from 0.2 to 10 kHz. BTE microphones: left front (LF), left rear (LR), right front (RF) and right rear (RR). CIC microphones: left (L) and right (R).

close to zero. The HRTFs measured in the condition HAD showed an average SD of 0.46 dB, which indicates less reliability than that found in the condition Control, however still providing reasonable reliability. Even in the condition HAD, the HRTFs were quite similar, however frequency dependent variability was observed, with larger deviations at higher frequencies. The condition Tripod1 was similar to the condition HAD, with an average SD of 0.34 dB. Also in the condition Tripod1, the deviations were frequency dependent, with larger variability at higher frequencies. In addition, larger deviations were also observed at lower frequencies.

The similar reliability in the conditions HAD and Tripod1 indicates that the exact position of the printed head was less important than the position of the hearing aids. It therefore appears that the variability in the Tripod1 condition was caused mainly by small errors in re-positioning the hearing aids. The larger variability between measurements at higher frequencies seems reasonable, since both conditions, HAD and Tripod1 were subject to small changes in hearing-aid position. The condition Tripod1 also revealed variability at lower frequencies (below 1 kHz), which indicates that the re-positioning of the head and torso involved larger position changes.

Table I shows SDs for four microphones of the BTE hearing aids (2 microphones per ear) and for both microphones of the CIC hearing aids (1 microphone per ear). Note that the conditions HAD was not measured for the CIC hearing aid because of difficulties with repositioning the hearing aid. Table I shows that the SDs were similar across microphones for each condition indicating that the variability does not depend on the type of the hearing aid. Notice that the CIC hearing aid has a SD comparable to the BTE for the Tripod case even though the hearing aid was not repositioned. The CIC measurements might be more sensitive towards movements of the printed head, because pinna effects arise due to the hearing aid position in the ear canal.

3.2. Across-session repetition

Figure 3 compares two HRTFs across two measurement sessions for both the printed head and the hu-

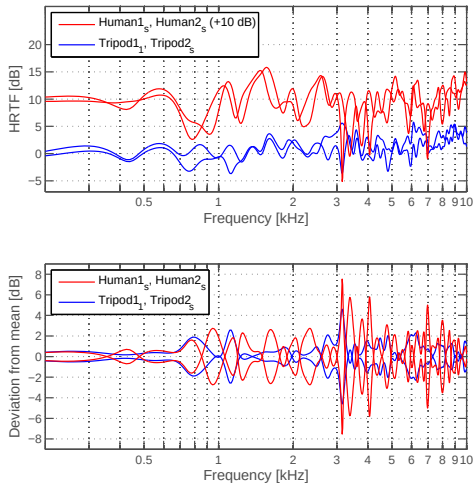


Figure 3. HRTF magnitude spectra for elevation 15° and measurements $Tripod1_1$ and $Tripod2_s$, and $Human1_s$ and $Human2_s$. Top panel: Absolute magnitudes with an offset of 10 dB between conditions for visualization purposes. Bottom panel: Deviation from the average. Notice the mirrored curves, which occur because the mean is calculated from only two samples.

man subject. The measurements used are the first Tripod1 measurement, $Tripod1_1$, $Tripod2_s$, $Human1_s$ and $Human2_s$. The average SDs were 0.88 and 1.5 dB for the two conditions, respectively. The variation between measurements from different sessions are clearly larger than that from the same session. Between the sessions, changes in the measurement setup such as re-positioning of acoustic damping material on floor and equipment, re-wiring of the hearing aids, as well as small temperature changes might have been involved. The variation in our HRTFs measured at different days is comparable to the variations shown in [7] measured under similar conditions.

A larger SD was obtained for the $Human1_s$ and $Human2_s$ measurements, as compared to the SD obtained for the $Tripod1_1$ and $Tripod2_s$. This corresponds to a larger variability when HRTFs are measured for the actual human listener than when measured using the printed head. The higher variability is most likely the result of small but unavoidable movements of the head and body, for instance due to breathing. These issues are not involved in measurements performed on the printed head, being reflected in smaller SD and thus better reliability of the measurements using the printed head.

4. Conclusions

Repeating HRTF measurements yields a variability in the results. Simple repetition without any repositioning of the microphones were very reliable, showing standard deviations about 0.01 dB. Repositioning of the hearing aids had a larger effect on the reliability, with standard deviations up to 0.6 dB. Additional repositioning of the printed head did not substantially increase the variability. HRTFs measured in different sessions, however, showed a larger variability attributed to less controlled measurement conditions. Finally, measurements with an actual human listener showed more variability than those with a printed head, most probably due to small head or body movements, which are unavoidable when dealing with a human listener. Our results indicates that the reliability of HRTF measurements with human listeners has its limitations, but that more reliable HRTFs can be measured when using a head simulator instead.

References

[1] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, Z. Tang: Approximating the head-related transfer function using simple geometric models of the head and torso. The Journal of the Acoustical Society of America **112** (2002) 2053–2064.

[2] K. Watanabe, S. Takane, Y. Suzuki: A novel interpolation method of hrtfs based on the common-acoustical-pole and zero model. Acta acustica united with acustica **91** (2005) 958–966.

[3] F. L. Wightman, D. J. Kistler: Factors affecting the relative salience of sound localization cues. Binaural and spatial hearing in real and virtual environments **1** (1997) 1–23.

[4] J. Fels, P. Buthmann, M. Vorlander: Head-related transfer functions of children. Acta Acustica united with Acustica **90** (2004) 918–927.

[5] J. C. Middlebrooks: Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. The Journal of the Acoustical Society of America **106** (1999) 1493–1510.

[6] S. Harder, R. R. Paulsen, M. Larsen, S. r. Laugesen, M. Mihocic, P. Majdak: A pipeline of geometry acquisition, 3-d printing, simulation, and measurement of head-related transfer functions. Submitted.

[7] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen: Head-related transfer functions of human subjects. Journal of the Audio Engineering Society **43** (1995) 300–321.

[8] F. L. Wightman, D. J. Kistler: Headphone simulation of free-field listening. i: Stimulus synthesis. The Journal of the Acoustical Society of America **85** (1989) 858–867.

[9] S. Harder, R. R. Paulsen, M. Larsen et al.: A three dimensional children head database for acoustical research and development. Proceedings of Meetings on Acoustics, 2013, Acoustical Society of America, 050013.

[10] P. Majdak, P. Balazs, B. Laback: Multiple exponential sweep method for fast measurement of head-related transfer functions. Journal of the Audio Engineering Society **55** (2007) 623–637.

[11] J. M. Kates: Digital hearing aids. Plural Pub., 2008.

Appendix D

Optimizing hearing-aid directionality from measurements and simulations

Optimizing hearing-aid directionality from measurements and simulations

S. Harder,[1, a)] R. R. Paulsen,[1] M. Larsen,[2] M. S. Pedersen,[2] S. Laugesen,[3] M. Mihocic,[4] and P. Majdak[4]
[1)] *Technical University of Denmark, DTU Compute, Department of Applied Mathematics and Computer Science, DK- 2800 Lyngby, Denmark.*
[2)] *Oticon A/S, DK-2765 Smørum, Denmark.*
[3)] *Eriksholm Research Centre, DK-3070 Snekkersten, Denmark.*
[4)] *Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria.*

(Dated: 9 March 2015)

We present a comparison of front-omni, hyper-cardioid, HATS-optimized (Brüel & Kjær head and torso simulator) and directional filters individually optimized from head related transfer functions (HRTFs), all measured on a behind-the-ear (BTE) hearing aid for four human test subjects. Furthermore, an optimal filter is calculated using finite element method (FEM) simulated HRTFs based on a listener-specific head model for one of the test subjects. We show that the simulated optimal filter results in a high Articulation-Index weighted Directivity Index (AI-DI) value, however the improvement compared to using the non-individualized filters, hyper-cardioid and HATS, is limited in this example. Based on AI-DI values for all four test subjects, found with the hyper-cardioid and HATS filters, we suggest that some individuals might have a larger benefit from a simulated listener-specific directional filter than seen in the aforementioned example.

PACS numbers: Valid PACS appear here
Keywords: Directivity, BTE hearing aid optimization, listener-specific directionality, acoustic measurements, acoustic simulations

I. INTRODUCTION

The primary purpose of a hearing aid is to provide amplification to make soft sounds audible to the hearing impaired user. However, a classical hearing impaired individual (sensorineural hearing loss) has more deficits than a lower sensitivity to sounds. These deficits include a lower dynamic range, a reduced frequency resolution, and a reduced temporal resolution. These are all issues that reduce speech intelligibility, even though a compressor to some extend can compensate for a lower dynamic range and a reduced temporal resolution.

Dillon[1] states with the benefits of a modern well fitted hearing aid, only two methods have proven to increasing intelligibility further. The first method is move the microphone closer to the source and the other method is to suppress off-target sound using a directional microphone. Since it is not always practical to move the microphone closer to the source, the directional microphone is of great importance for improving intelligibility in modern hearing aids.

A directional microphone works by combining the acoustical input from two omni-directional microphones, using a set of directional filters. The performance of a directional microphone is often measured in terms of the AI-DI (Articulation-Index weighted Directivity Index). The theoretical maximum AI-DI of a standard hearing aid directional microphone is 6 dB, and in critical listening conditions an additional 1 dB of AI-DI corresponds to about 10 % better speech understanding[1].

Today, the directional filters are designed once for each hearing-aid model, with the objective of obtaining the best possible AI-DI average across the population. Such an average filter is usually obtained by optimizing the directional filters with respect to AI-DI using head related transfer functions (HRTFs) measured on a head and torso simulator, such as the Brüel & Kjær HATS. It is, however, known that the directional microphone performance varies considerably across the population. Valente et al.[2] reported a mean advantage in signal-to-noise ratio (loudspeaker at 0 degree, noise at 180 degree azimuth) of 7.4 dB using a directional microphone instead of an omni-directional microphone. The individual SNRs were ranging from 3.5 dB to 16.1 dB, with a standard deviation of 3.0 dB. The varying individual gain from using directional microphones suggests that the HATS-optimized directional filters might be less than optimal for the individual hearing-aid user.

The solution to obtaining optimal directionality for the individual hearing-aid user is, in principle, straightforward. Acoustical measurements taken from a large number of loudspeakers positioned on a sphere around the hearing-aid user can be used for individual optimization of the directional filters[3,4]. Such measurements are, however, very cumbersome and completely intractable in clinical practice.

It has been proposed[3] to replace the aforementioned elaborate acoustical measurements with numerical simulations, for instance based on the FEM (Finite Element Method). This approach requires an accurate 3D geometrical representation of the individual hearing-aid user's head. Such an individual 3D head model can be obtained using for instance the surface scanning, stitching and meshing procedure proposed by Harder et al.[5]. It is however intractable to acquire an expensive 3D surface scanner in a typical clinic. The possibility of obtaining

[a)]Electronic mail: sthar@dtu.dk

3D head models from simple 2D images is therefore attractive. Estimation of 3D structures based on 2D images is an on-going research theme in the field of multiple-view geometry[6] where highly calibrated camera setups are normally used. Alternatively, strong statistical priors can be used to predict structures with known statistical shape priors from one or a few 2D projections. In particular, it has proven possible to estimate the 3D anatomy of human faces from frontal photos[7]. This approach can potentially be extended to predict the entire head shape.

In this paper we examine part of the question of interest: Given an accurate listener-specific 3D head model, is it possible to obtain maximal directivity using FEM simulations?

We compute the maximal listener-specific directivity using HRTFs measured on a behind-the-ear (BTE) hearing aid, and we investigate how close to the maximal directivity one can get using simulated data for the optimization. Optimal listener-specific directional filters are calculated using state-of-the-art measurements of listener-specific HRTFs. Furthermore, we use a listener-specific 3D head model for FEM simulations of HRTFs. The simulated HRTFs are applied to calculate optimal directional filters, and the two cases are compared using DI-curves and AI-DI values.

II. MATERIAL AND METHODS

A. Acoustical Measurements

The acoustical measurements employed in this paper were carried out using the same method and measurement setup that was described in Harder et al.[8]. In Harder et al.[8] a set of HRTFs were measured for a microphone placed in the ear canal. Here we measure HRTFs for microphones build into a BTE hearing-aid shell. Data measured with the same setup has furthermore been published in Harder et al.[9], where repeated HRTF measurements for microphones in a BTE shell were analyzed using frontal source positions (azimuth angle = 0 °). For completeness, we here provide a brief summary of the measurement procedure. The keen reader can find more details in Harder et al.[8].

The HRTFs were measured in a semi-anechoic chamber. Twenty-two loudspeakers were mounted on an arc at fixed elevations ranging from −30 ° to 80 °. All equipment was covered with acoustic damping material. The human test subject was placed in a chair in the center of the arc. The human subject was wearing BTE shells with microphones and connecters on both ears. An exponential frequency sweep beginning at 50 Hz and ending at 20 kHz was used to measure each HRTF.

A tracking sensor was mounted on top of the listeners head to assure that head movements were within a set of valid ranges. The ranges were set to ±2.5 cm for position, ±2.5 ° for azimuth, and ±5 ° for elevation and roll. A complete measurement session lasted for approximately 20 minutes and a total of 1550 HRTFs were measured for each listener.

For this particular study HRTF of four listeners (NH166, NH167, NH168 and NH170) were measured. The naming convention follow the ARI database `http://sofaconventions.org`, where all data is available.

B. Post-processing

The purpose of the post-processing method is to remove room and loudspeaker responses from the HRTF data, with a focus on maintaining the small phase differences that exist between the BTE microphones. This method was presented in Harder et al.[9].

For each raw HRTF, head-related impulse responses (HRIRs), i.e., the inverse Fourier transform, was calculated, yielding the raw HRIRs. The raw HRTFs were affected by both the loudspeaker, room and microphone used for the measurements. These effects can be described by the equipment transfer functions (ETFs), which were derived from a reference measurement in which the BTE microphones were placed in the center of the arc and the impulse response was measured for all loudspeakers. A set of ETFs were measured each time before the HRTF measurement of a subject.

For a given microphone k, a compensated HRTF $C_k(\theta, \phi)$ at elevation θ, and azimuth ϕ, was calculated in the frequency domain:

$$C_k(\theta, \phi) = \frac{H_k(\theta, \phi)}{E_F(\theta)}, \tag{1}$$

where $H_k(\theta)$ is the windowed raw HRTF for microphone k and $E_F(\theta)$ is the windowed ETF for the front microphone in the BTE hearing aid.

In this way the HRTFs have been compensated for room and loudspeaker responses, however they are not compensated for the individual microphone response. The small phase differences that exist between the BTE microphones are highly important with respect to directivity. We learned that the best way to retain the phase difference during the post processing was to use the ETF for the same microphone for a set of BTE microphones, instead of compensating each microphone using their respective EFT.

C. FEM Simulations

A listener-specific head and ear model of test subject 167 was required for the FEM simulations. A framework for generating listener-specific head models was presented in Harder et al.[5], and in Harder et al.[8] we used a listener-specific head model of subject 167 for FEM simulations of a microphone placed in the ear canal. Here we use the same head and ear model for FEM simulations of the microphones in a BTE hearing aid.

Simulating HRTFs for a BTE hearing aid placed on top of the ear requires two tasks; placement of the BTE hearing aid including a deflection of the pinna, and a harmonical simulation resulting in HRTFs. Both simulations were performed in ANSYS 15.0.

1. BTE placement

The position of the hearing aid when worn is determined by its own shape and by the shape of the ear. In the physical world, the hearing aid is placed on the ear by pressing it down and letting go. Then the ear will deflect and hold the hearing aid in place, and the hearing aid's final position will be determined by the forces and friction between the ear and hearing aid.

A similar procedure was used in a static structural simulation: the hearing aid was forced on the ear by a displacement, and then released. As in the physical world, when the hearing aid is released, the ear deforms, and the hearing aid position depends on the contact between ear and instrument, and on the stiffness of the ear.

The BTE placement was based on the 3D head model from Harder et al.[8] and a 3D geometry of a standard BTE shell. To save calculation time, the ear was cut from the rest of the head model. A mesh was generated for the simulation with 25,580 nodes for the ear and 4,754 nodes for the hearing aid. The ear model was constrained with fixed nodes on all the cut surfaces.

Our calculations consisted of three sub-steps, see Figure 1. In Figure 1a there is no displacement yet, however the BTE hearing aid and ear are in contact. In Figure 1b the BTE has a maximal displacement and in Figure 1c the BTE has been released. We used a stiffness with 4 MPa Young's modul and 0.4 Poisson's ratio, which corresponds to a soft silicone material, and the contact was defined as 'frictional' with a coefficient of 0.5.

After the simulation, the deformed mesh of the ear and hearing aid was saved in Initial Graphics Exchange Specification (IGES) format and then re-imported into the head, replacing the undeformed ear and hearing aid in it's original position. The updated geometry was then used for the simulation.

2. Acoustical simulations

The acoustical simulations were based on the head geometry obtained in Harder et al.[8] with the BTE placement and pinna bending calculated in Section II C 1. The head was mounted on a virtual model of the Brüel & Kjær torso simulator. An air-filled box with dimensions 420 mm×700 mm×250 mm was created around the head and torso, and the model was subtracted from the box. The outcome is the air surrounding the model. A 40 mm perfectly matched layer (PML) was added to the outside of the inner air-box. The PML assures that sound radiated from the inner air part are not reflected back. The default setting for attenuation in the PML region was used (0.001 or −60 dB), yielding 120 dB total attenuation of the sound field reflected from the PML.

The head geometry was meshed with first-order acoustical elements (ANSYS type 30). In order to reduce calculation times two meshes were created: one for frequencies below 7.5 kHz and one for frequencies up to 10 kHz. Both meshes were refined at the ears to an average edge length of 1 mm. The 7.5-kHz mesh had a length of 6 mm for the rest of the model, whereas the 10-kHz mesh used a length of 4 mm for the rest of the model. The re-meshing method was 'Hex dominant' for both meshes. A hexahedral mesh was preferred to a tetrahedral mesh, since it produces a mesh with less nodes. The 7.5-kHz mesh contained 1.1 million nodes, whereas the 10-kHz mesh contained 2.3 million nodes.

In the acoustical measurement setup (see Section II A), loudspeakers surrounding the head generated the sound and the microphones captured the sound pressure. In the simulation, the reciprocal approach was used[10,11]: the sound source was placed at the microphone position, and the sound pressure level was calculated on a sphere surrounding the head. The reciprocal approach yields idential results to the direct approach, but the simulation can be performed for all sound-source directions in a single simulation step[12].

The simulation was performed for a total of 54 linearly spaced frequencies from 187.5 to 10,125 Hz and for the front and rear microphones incorporated into the BTE shell on the left ear. The resulting far field was evaluated in vertical and horizontal angle steps of 2.5 degrees at a distance of 1.2 m. However, in this paper we only use 1550 source positions, equivalent to the experimental setup used for measuring HRTFs.

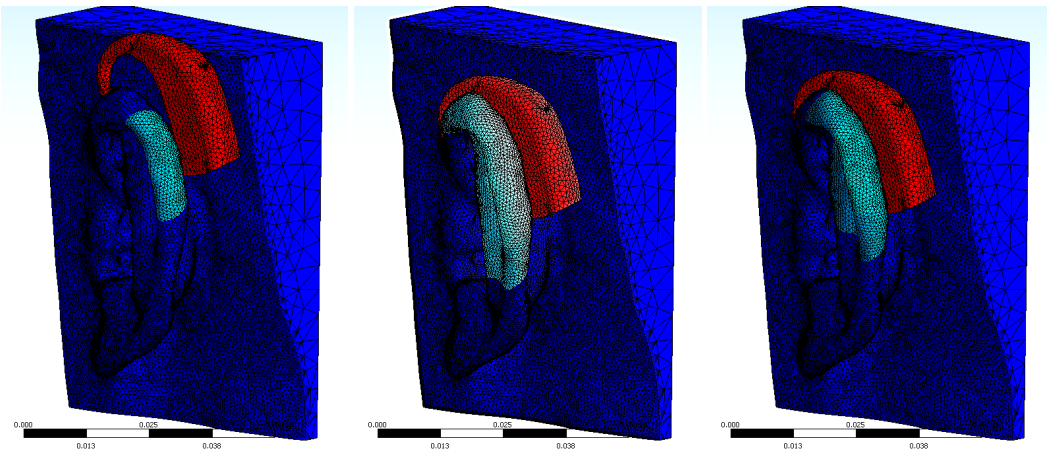
Furthermore, a free-field transfer function was simulated and used as a reference for the other simulations, resulting in a set of simulated HRTFs comparable to the measured HRTFs described in Section II A. In this way both measured and simulated HRTFs were ratios of sound pressures at a microphone on the head and at a microphone in the center of the head (head absent).

D. 3D DI method

The directivity index (DI) is a measure of how much sound is received from the frontal position, zero elevation and azimuth angle, compared to how much sound is received from all other directions. Equation 2 contains a mathematical description of the DI.

$$DI(f) = 10\log_{10}\left[\frac{4\pi|p(0,0)|^2}{\int_0^{2\pi}\int_0^{\pi}|p(\theta,\phi)|^2\sin(\theta)d\theta d\phi}\right], \quad (2)$$

where $|p(0,0)|^2$ is the magnitude-squared on-axis sound-pressure microphone response to a plane wave and



(a) Initial position. (b) Maximal displacement. (c) Final position.

FIG. 1: The three step BTE placement. A displacement is applied to the hearing aid at the initial position, whereupon the hearing aid is released. The final position is determined by the shape of the ear and hearing aid, the stiffness of the ear and the frictional forces between ear and hearing aid.

$|p(\theta,\phi)|^2$ is the spatially averaged magnitude-squared sound pressure microphone response.

As explained in Dittberner[13], the on-axis free-field response of a microphone is easily computed. However, the spatially averaged response is not easily computed. We utilize the free-field method described by Dittberner to calculate the spatially averaged response[13]:

$$DI(f) = 10\log_{10}\left[\frac{4\pi\cdot|p(0,0)|^2}{\sum_{m=1}^{\frac{2\pi}{\Delta\phi}}\sum_{n=1}^{\frac{\pi}{\Delta\theta}}|p(\theta_n,\phi_m)|^2\sin(\theta_n)\Delta\theta\Delta\phi}\right], \tag{3}$$

where the spatially averaged response is estimated from a discrete number of source positions that should be evenly distributed on a sphere surrounding the head.

According to Equation 3 it is of great importance when calculating DI that the source positions are evenly distributed on the sphere surrounding the head. In our case, we have measurement data where the source positions are not evenly distributed on a sphere, see Section II A. Source positions are missing for elevation angles below $-30\,^\circ$ and the frontal part is more densely sampled.

We developed a weighting scheme, inspired by the method by Dittberner et al.[13], that accounts for an uneven distribution of source positions. In Dittberner et al.[13] the number of azimuth angles per elevation is calculated based on the surface area that each elevation angle covers of the entire sphere. Here, the basic idea is to calculate the surface area covered by a particular source position and weigh the measurement for that position accordingly. Positions with a large surface area are given a large weight, whereas smaller surface areas are given a smaller weight.

The surface area of a sphere slice covered by the ith elevation θ_i is:

$$S_E(\theta_i) = \int_{l_i}^{u_i} 2\pi g(z)\sqrt{1+[g'(z)]^2}dz, \tag{4}$$

where $g=\sqrt{1-z^2}$, $u_i = \theta_i + \left(\frac{\theta_{(i+1)}-\theta_i}{2}\right)$ and $l_i = \theta_i - \left(\frac{\theta_i-\theta_{(i-1)}}{2}\right)$. Here it is assumed that g revolves around the z-axis and that the radius of the sphere is one (the radius is insignificant, since only ratios are considered). The boundaries u_i and l_i are positions on the z-axis and are calculated as the sine of the angular-boundaries of a given elevation.

The surface area, S_{EA}, for the ith elevation and jth azimuth angle is now given by:

$$S_{EA}(\theta_i,\phi_j) = \frac{(\phi_{j+1}-\phi_{j-1})/2}{2\pi}\cdot S_E(\theta_i). \tag{5}$$

Each measurement point is weighed according to the size of the surface area that it covers, using the weights calculated in Equation 6:

$$\alpha(\theta,\phi) = \frac{S_{EA}(\theta,\phi)}{4\pi}. \tag{6}$$

The final computation of DI, as a function of frequency, is:

$$DI(f) = \frac{4\pi|C(0,0)|^2}{\sum_{m=1}^{M}|C(\theta_m,\phi_m)|^2 \cdot \alpha(\theta_m,\phi_m)}, \quad (7)$$

where (θ_m, ϕ_m) is the mth source position and where M is the total number of source positions.

E. Optimization of DI

Equation 7 provides a DI value calculated for a single set of HRTFs. One can however also combine two or more omni-directional microphones to achieve a directional microphone. Here a BTE hearing aid, which combines two omni-directional microphones, is of interest. The response from such a directional-microphone becomes:

$$C_D(\theta,\phi) = C_F(\theta,\phi)w_1 + C_R(\theta,\phi)w_2, \quad (8)$$

where $C_F(\theta,\phi)$ and $C_R(\theta,\phi)$ are the HRTFs for the front and rear microphone, respectively, and w_1 and w_2 are complex weights.

Using matrix notation:

$$\mathbf{C}_D = \mathbf{C}\mathbf{w}, \quad (9)$$

where $\mathbf{C} = \begin{bmatrix} C_F(0,0) & C_R(0,0) \\ \vdots & \vdots \\ C_F(\theta_m,\phi_m) & C_R(\theta_m,\phi_m) \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$.

We are optimizing Equation 7 using a directional microphone as defined in Equation 8 and 9:

$$\begin{aligned} DI(f) &= \frac{4\pi|C_D(0,0)|^2}{\sum_{m=1}^{M}|C_D(\theta_m,\phi_m)|^2 \cdot \alpha(\theta_m,\phi_m)} \\ &\propto \frac{(\mathbf{C}_0\mathbf{w})^H(\mathbf{C}_0\mathbf{w})}{(\mathbf{C}_\alpha\mathbf{w})^H(\mathbf{C}_\alpha\mathbf{w})}, \end{aligned} \quad (10)$$

where ^{H} denotes the complex conjugate transpose,

$$\mathbf{C}_0 = [C_F(0,0)\ C_R(0,0)] \quad (11)$$

and

$$\mathbf{C}_\alpha = \begin{bmatrix} C_F(0,0)\sqrt{\alpha(0,0)} & C_R(0,0)\sqrt{\alpha(0,0)} \\ \vdots & \vdots \\ C_F(\theta_m,\phi_m)\sqrt{\alpha(\theta_m,\phi_m)} & C_R(\theta_m,\phi_m)\sqrt{\alpha(\theta_m,\phi_m)} \end{bmatrix}. \quad (12)$$

The cost function in Equation 10 can be optimized using the generalized eigenvalue method as suggested in Gay et al.[14]. Here $\mathbf{w}$ is found as the eigenvector corresponding to the maximum generalized eigenvalue. We use the MATLAB function `eig` in our optimization procedure.

1. Evaluation on free-field model

The DI method described in Equation 7 was evaluated using a set of synthesized data representing two microphones in free field placed 10 mm apart, resembling a BTE hearing aid. The front microphone was placed in the center of the coordinate system and the rear microphone was placed 10 mm away along the negative x-direction. Equation 13 was used to synthesize the free-field HRTFs, C_{FF}, for each frequency, f, and source position, (θ,ϕ), individually:

$$C_{FF}(f,\theta,\phi) = \exp\left(\frac{-i2\pi f d\cos(\rho(\theta,\phi))}{c}\right), \quad (13)$$

where d is the distance from the center of the coordinate system along the x-axis, $\rho(\theta,\phi)$ is the angle between the x-axis and a given source position (θ,ϕ).

Two sets of source positions were evaluated. The first set consisted of the 1550 source positions that was used for the measurements. The second set consisted of the 1550 source positions plus an additional 448 source positions to cover the missing data on the lower part of the sphere, resulting in 1998 positions. The two set of HRTFs are from now on referred to as Syn_{1550} and Syn_{1998}.

Figure 2 left and right show DI as a function of frequency for Syn_{1998} and Syn_{1550}, respectively. Front-omni, cardioid, and optimal filters according to Equation 10 to 12 are used to calculate DI curves. The cardioid and optimal cases both show a roll-off above 10 kHz, which arise because the wavelength becomes comparable to the distance between the microphones. It is reassuring to observe that our optimal DI value is equal to 6 dB corresponding to the theoretically optimal hypercardioid[1].

It is common practice to average over frequencies to obtain a single DI value[13]. However, it is not the entire frequency range that is important for speech intelligibility. The articulation index (AI)[13] is a frequency weighting scheme where frequencies that are important for speech is weighted high. We calculate an articulation index directivity index (AI-DI) for the cases omni, hyper-cardioid and optimal for both Syn_{1998} and Syn_{1550}. The values are seen in Table I. The table also contain 'asymptotical' DI (AS-DI) values, which are the values that the DI-curves approximate when frequencies tend towards zero, see Figure 2.

The evaluation of our 3D DI surface-area weighting method reveals that the theoretically correct AS-DI values were obtained for Syn_{1998}, where points on a complete sphere were available. For Syn_{1550}, where points were missing in the lower part of the sphere the AS-DI values dropped 0.2 and 0.4 dB for the hyper-cardioid and optimal filers, respectively. The error can be contributed to the missing source positions in the lower part of the sphere.

AI-DI values are calculated for Syn_{1998} and Syn_{1550}, which show a maximal drop of 0.2 dB compared to the

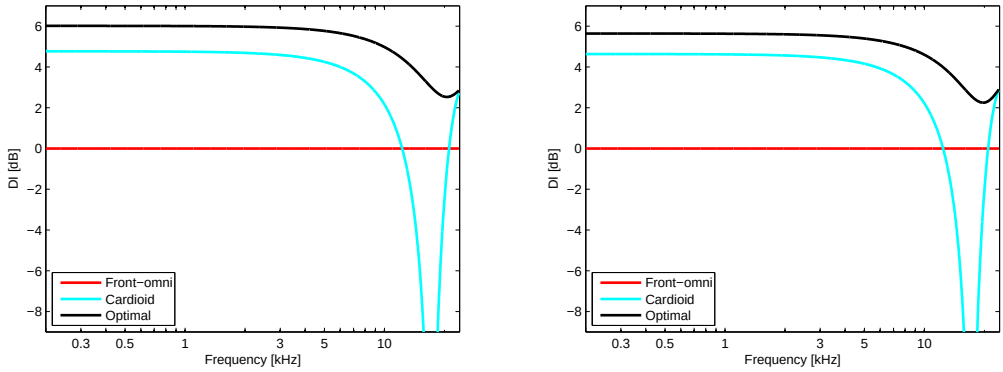


FIG. 2: DI-curves calculated based on free-field HRTFs. Left: Syn_{1998}, Right: Syn_{1550}.

		Front-omni	Cardioid	Optimal
AS-DI	Theoretical	0.0 dB	4.8 dB	6.0 dB
	Syn_{1998}	0.0 dB	4.8 dB	6.0 dB
	Syn_{1550}	0.0 dB	4.6 dB	5.6 dB
AI-DI	Syn_{1998}	0.0 dB	4.6 dB	5.9 dB
	Syn_{1550}	0.0 dB	4.5 dB	5.6 dB
	Unw_{1550}	0.0 dB	3.6 dB	4.4 dB

TABLE I: AI-DI and asymptotical DI (AS-DI) values. Theoretical values are obtained from[1], whereas the remaining values are calculated based on synthesized free-field HRTFs for a BTE. Unw_{1550} show the unweighted Syn_{1550}.

AS-DI values. These deviations are solely due to the roll-off caused by microphone distance.

Unweighted AI-DI values, Unw_{1550}, are also calculated for Syn_{1550}. They show a dramatic drop in AI-DI value of 0.9 and 1.2 dB compared to the Syn_{1550}, which has been weighted according to our DI weighting scheme. This shows that our DI weighting scheme, to a great extend, compensates for the missing source positions, even when source positions are missing on a large region of the sphere.

III. RESULTS

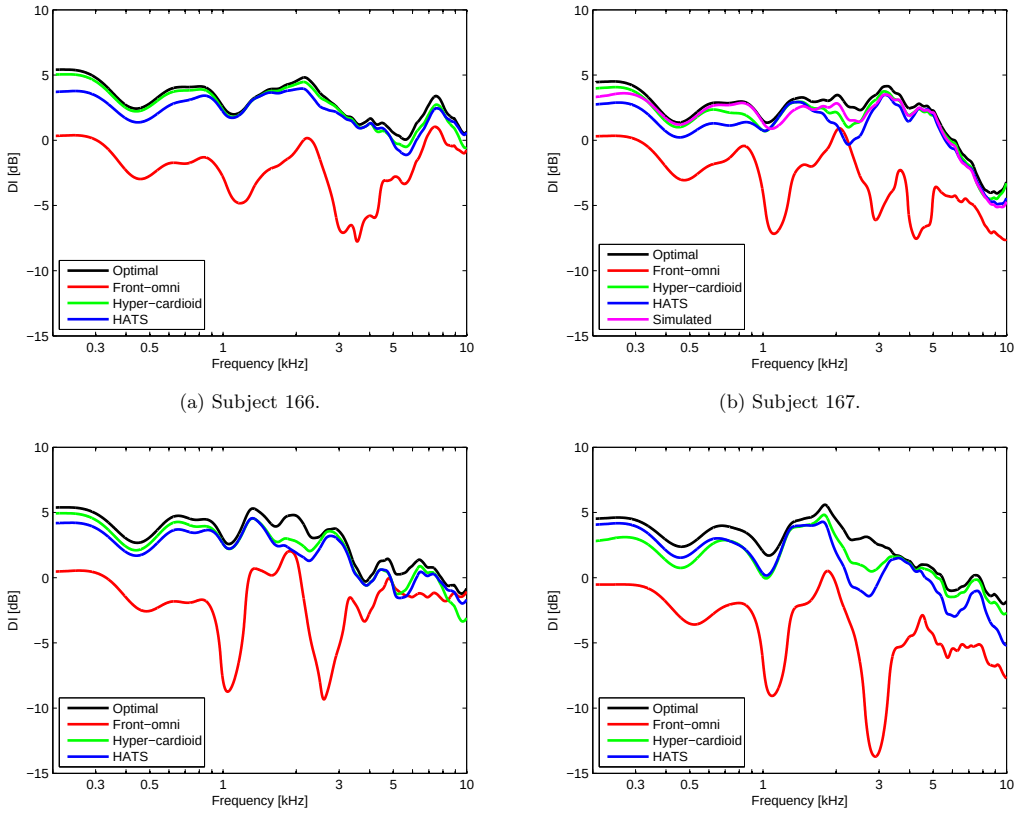
The major goal of this paper is to investigate if maximum directionality can be approximated by optimizing filters using HRTFs simulated from a listener-specific head model. Figure 3 contains DI as a function of frequency for optimal, front-omni, hyper-cardioid and HATS-optimized filters. All four results are from the left ear. The HATS filter is obtained using HRTF measurements on a HATS (Brüel & Kjær) head and torso simulator in the horizontal plane. The HATS filter is calculated as a mean of the filters obtained from the left and the right ear measurements. The optimal filters are calculated using the method described in Section II E. For test subject 167 filters optimized from a set of simulated HRTFs are included. The curves are all logarithmically smoothed with a 1/3-octave wide rectangular window to obtain patterns which can be compared more easily.

Figure 3 reveals first of all that the omni-directional microphone has a very fluctuating DI-curve. At some frequencies the omni has DI values as low as −5 to −10 dB, which are seen as large dips in the DI-curves at 1 and 3 kHz. These dips probably arise due to reflections from torso, thighs, or equipment.

Furthermore the figure reveals that the optimized DI-curve is frequency dependent, with a lower directionality at high frequencies (above 5 kHz). This tendency is less clear for subject NH166. For low frequencies (below 1 kHz) the gain in directionality between front-omni and optimal filters are relatively constant between 4 to 6 dB, for the different individuals.

The hyper-cardioid and HATS-optimized filters have a varying performance for the different individuals. The hyper-cardioid performs almost as good as the optimal filter for subject NH166 and NH167, however the performance decreases a bit for subject NH168 and even more for subject NH170. The HATS filters are in general worse than the hyper-cardioid filters. The HATS-filter curve has a tendency to decrease for low frequencies, which could be caused by a mismatch between the microphone pairs used for the human and the HATS measurements. The mismatch tendency is not seen for subject NH170, in this case indicating that the microphone pairs are possibly better matched.

It is surprising that our data indicates that a theoretical hyper-cardioid filter provides higher directivity than a HATS-optimized filter. However, our HATS filters are based on two measurements, whereas HATS filters in a commercial hearing aid would be an average over several



(a) Subject 166.

(b) Subject 167.

(c) Subject 168.

(d) Subject 170.

FIG. 3: DI-curves calculated based on measured HRTFs on the left ear. Optimal, front-omni, hyper-cardioid, and HATS-optimized filters are applied. For test subject NH167 filters optimized on simulated HRTFs are included.

measurements. This would probably provide a filter that is more robust towards for example hearing aid position. Furthermore, our HATS filter is measured using a single microphone pair, which probably to some degree suffer from a microphone mismatch. The microphone pair used for commercial HATS filter measurements, would probably be chosen so that they match very well, or the filters could be an average of several microphone pairs evening out microphone mismatches.

Test subject NH167 is also evaluated using filters optimized on simulated HRTFs. The DI-curve is seen in Figure 3b. The performance of the filters for the simulated data is quite high, however comparable to the performance obtained for the hyper-cardioid or HATS-optimized filters.

Table II sums up the frequency dependent DI-values in terms of AI-DI values. Besides the data already presented for the four test subjects, Table II also contains a remeasurement of subject NH167, named NH167-2. Furthermore, data from the right ears have been included. The table reveals an increase in AI-DI value of 0.6 and 0.3 dB for the two NH167 measurements, when the filters are changed from HATS-optimized to simulated. The difference between hyper-cardioid and simulated AI-DI values is however 0.1 and −0.1 dB, suggesting no improvement in directivity from using individually simulated filters. However, looking at the other test subjects, it is noticed that the hyper-cardioid and HATS-optimized filters perform very well for some individuals and worse for others. Subject NH167 has a very high AI-DI with the hyper-cardioid and also with the HATS-optimized filters, which means that there is very little room for improving the AI-DI using simulated filters.

The inter-individual variation for the optimal AI-DI values in the second column in Table II is also worth noticing. The AI-DI values have a mean of 3.2 dB and a

standard deviation of 0.4 dB. It is striking that one individual can achieve an AI-DI value of 4.1 dB while another individual has to settle with 2.6 dB. Furthermore, notice that measurement NH167-2 has an inter-ear difference of 0.9 dB using optimal filters, while the inter-ear difference is 0.0 dB for measurement NH167 even though the two measurements are performed on the same individual. This finding suggests that the AI-DI values are dependent on other aspects than head and ear shape. We believe that differences in hearing-aid placement between the measurement series is the most likely explanation for these deviations.

			Relative to F.O.			
	F.O.	Opt.	Opt.	H.C.	HATS	Sim.
NH166 L	-2.1	3.3	5.4	5.1	4.7	-
NH166 R	-1.8	3.1	4.9	4.3	4.0	-
NH167 L	-2.0	2.8	4.8	4.2	3.7	4.3
NH167 R	-2.4	2.8	5.2	4.7	4.3	-
NH167-2 L	-2.4	2.6	5.0	4.6	4.2	4.5
NH167-2 R	-2.1	3.5	5.6	5.3	5.0	-
NH168 L	-1.1	3.6	4.7	4.0	3.6	-
NH168 R	-2.0	3.1	5.1	4.4	3.9	-
NH170 L	-2.9	3.2	6.1	5.0	4.9	-
NH170 R	-1.3	4.1	5.4	5.1	4.8	-
NH169 L	-1.9	3.2	5.1	4.8	4.4	4.8
NH169 R	-2.0	3.3	5.3	5.1	4.7	-
Mean	-2.0	3.2	5.2	4.7	4.4	-
SD	0.5	0.4	0.4	0.4	0.5	-

TABLE II: AI-DI values calculated from measured HRTFs. Optimal (Opt.), hyper-cardioid (H.C.), HATS, and simulated (Sim.) values are given relative to front-omni. Mean and standard deviations (SD) are also provided for most filters.

A. Printed head model

We previously presented a method for printing a 3D head from a listener-specific head model[8]. The 3D printed head was based on the 3D head model used for the FEM simulations in Section II C. The 3D printed head was mounted on a HATS torso and was, similar to the human test subjects, equipped with a BTE shell and exposed to the acoustical measurements explained in Section II A. Front-omni, hyper-Cardioid, HATS-optimized, and simulated filters were applied to the HRTFs and a DI was calculated using the proposed weighting method. The DIs are shown as function of frequency in Figure 4. Notice that the simulated filter performs almost equal to the optimal filter. This suggests that the simulation has better agreement with HRTF measurements on the printed head than with measurements on the real human test subject. The better agreement could for instance arise due to more similar geometries (virtual model vs. human head/printed model) or more similar acoustic properties of material (virtual material vs. human skin/printing material). This suggests that our virtual head model has geometrical errors, and that a more accurate model would provide simulated AI-DI values which approached the optimal filters to a higher degree.

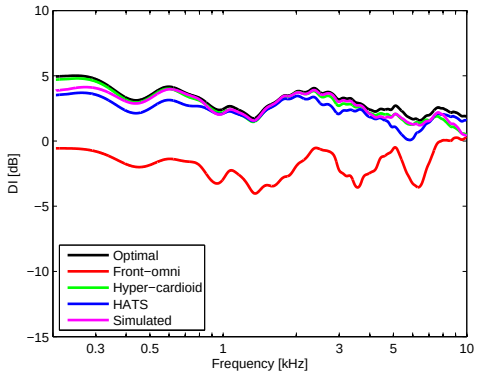


FIG. 4: DI-curves calculated based on left ear measured HRTFs for test subject NH169 (printed head model).

IV. DISCUSSION AND CONCLUSIONS

We showed that FEM simulations on a listener-specific 3D head model can provide directional filters that gives a relatively high AI-DI for the individual listener. Both AI-DI values and DI curves were calculated for front-omni, hyper-cardioid, HATS-optimized, and optimal filters. The gain obtained by changing from HATS or hyper-cardioid filters to simulated listener-specific filters was very limited or not existing for test subject NH167, since both HATS and hyper-cardioid filters perform very well.

We also presented optimal, front-omni, hyper-cardioid, HATS, and optimal AI-DI values for four human test subjects (eight ears) and DI curves for the left ear of each test subject. Our results suggests that there is a difference between the directivity obtained for the individual listener using the non-individualized filters. Because some individuals may benefit less from the non-individualized filters, they might have a larger benefit from having simulated individual filters than test subject NH167. The HATS filters, that we believe are somewhat similar to the filters used in commercial hearing aids, are measured for a head and torso simulator that represents an average of the human adult population. Since children have different head sizes and different proportions than an adult, we believe that children might benefit more from listener-specific simulated filters than adults. Other groups of

people who have head or ear geometries that are very different from the HATS geometry might also have larger benefit of a more listener-specific directional filter.

Acoustical measurements were also performed on a 3D printed version of the listener-specific head used for the FEM simulations. Both front-omni, hyper-cardioid, HATS, optimal and simulated optimal filters were applied and DI curves were generated. The result was that the hyper-cardioid and simulated optimal filters performed equally well and almost as well as the optimal filter. This indicates that the simulation are more similar to measurements on the 3D printed head than measurements on the human subject. We suggest that the larger difference between simulation and human measurements could arise due to small geometrical errors in the head model or due to differences in acoustical properties between human skin and virtual material properties in the simulation.

ACKNOWLEDGMENTS

The authors would like to thank the Oticon Foundation for funding this work.

[1]H. Dillon, *Hearing aids*, 2nd ed. (Boomerang Press, 2012).
[2]M. Valente, D. Fabry, and L. G. Potts, "Recognition of speech in noise with hearing aids using dual microphones," Journal of the American Academy of Audiology **6** (1995).
[3]K. B. Rasmussen, "Microphone system with directional response," (2007), uS Patent 7,212,642.
[4]S. Laugesen, K. B. Rasmussen, and T. Christiansen, "Design of a microphone array for headsets," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* (IEEE, 2003) pp. 37–40.
[5]S. Harder, R. R. Paulsen, M. Larsen, *et al.*, "A three dimensional children head database for acoustical research and development," in *Proceedings of Meetings on Acoustics*, Vol. 19 (Acoustical Society of America, 2013) p. 050013.
[6]R. Hartley and A. Zisserman, *Multiple view geometry in computer vision* (Cambridge university press, 2003).
[7]V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," Pattern Analysis and Machine Intelligence, IEEE Transactions on **25**, 1063–1074 (2003).
[8]S. Harder, R. R. Paulsen, M. Larsen, S. Laugesen, M. Mihocic, and P. Majdak, "A pipeline of geometry acquisition, 3-d printing, simulation, and measurement of head-related transfer functions,".
[9]S. Harder, R. R. Paulsen, M. Larsen, S. Laugesen, M. Mihocic, and P. Majdak, "A comparison between measured head related transfer functions - a repeatability study,".
[10]P. M. Morse, *Theoretical acoustics* (Princeton University Press, 1986).
[11]D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," The Journal of the Acoustical Society of America **120**, 2202–2215 (2006).
[12]F. Jacobsen and P. M. Juhl, *Fundamentals of General Linear Acoustics* (John Wiley & Sons, 2013).
[13]A. B. Dittberner and R. A. Bentler, "Predictive measures of directional benefit part 1: Estimating the directivity index on a manikin," Ear and hearing **28**, 26–45 (2007).
[14]S. L. Gay and J. Benesty, *Acoustic signal processing for telecommunication* (Kluwer Academic, 2000).
[15]P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," Journal of the Audio Engineering Society **55**, 623–637 (2007).
[16]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* (IEEE, 2001) pp. 99–102.
[17]B. Gardner, K. Martin, *et al.*, "Hrtf measurements of a kemar dummy-head microphone," Massachusetts Institute of Technology **280**, 1–7 (1994).

Bibliography

[ADTA01] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001.

[ARI] The ari database: `http://sofaconventions.org`.

[ASM+98] AF Ayoub, P Siebert, KF Moos, D Wray, C Urquhart, and TB Niblett. A vision-based three-dimensional capture system for maxillofacial assessment and surgical planning. *British Journal of Oral and Maxillofacial Surgery*, 36(5):353–357, 1998.

[BB02] Mark Burge and Wilhelm Burger. Ear biometrics. In *Biometrics*, pages 273–285. Springer, 2002.

[Bjå04] Jan G Bjålie. *Menneskets anatomi og fysiologi*. Gad, 2004.

[BL12] Seung-Yeob Baek and Kunwoo Lee. Parametric human body shape modeling framework for human-centered product design. *Computer-Aided Design*, 44(1):56–67, 2012.

[Bla83] Jens Blauert. Spatial hearing: the psychoacoustics of human sound localization, 1983.

[BV03] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, 2003.

[CB05] Hui Chen and Bir Bhanu. Contour matching for 3d ear recognition. In *Application of Computer Vision, 2005. WACV/MOTIONS'05*

Volume 1. Seventh IEEE Workshops on, volume 1, pages 123–128. IEEE, 2005.

[CB07] Hui Chen and Bir Bhanu. Human ear recognition in 3d. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):718–737, 2007.

[CBSV03] Kyong Chang, Kevin W Bowyer, Sudeep Sarkar, and Barnabas Victor. Comparison and combination of ear and face images in appearance-based biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1160–1165, 2003.

[CL96] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312. ACM, 1996.

[CW99] Corey I Cheng and Gregory H Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. In *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[DB07] Andrew B Dittberner and Ruth A Bentler. Predictive measures of directional benefit part 1: Estimating the directivity index on a manikin. *Ear and hearing*, 28(1):26–45, 2007.

[Dil12] Harvey Dillon. *Hearing aids*. Boomerang Press, second edition, 2012.

[DLT12] Gouri Dhatt, Emmanuel Lefrançois, and Gilbert Touzot. *Finite element method*. John Wiley & Sons, 2012.

[DM98] Richard O Duda and William L Martens. Range dependence of the response of a spherical head model. *The Journal of the Acoustical Society of America*, 104(5):3048–3058, 1998.

[eos] Sls printer eos: `www.eos.info`.

[EW05] C Elberling and K Worsøe. Når lyden bliver svagere. *Om hørelse og*, 2005.

[fab] The dtu fablab: `http://www.fablab.dtu.dk/`.

[fac] Facegen webpage: `www.facegen.com`.

[FB81] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[FDGR11] Zacharias Fourie, Janalt Damstra, Peter O Gerrits, and Yijin Ren. Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. *Forensic science international*, 207(1):127–134, 2011.

[FVDF+94] James D Foley, Andries Van Dam, Steven K Feiner, John F Hughes, and Richard L Phillips. *Introduction to computer graphics*, volume 55. Addison-Wesley Reading, 1994.

[GB00] Steven L Gay and Jacob Benesty. *Acoustic signal processing for telecommunication*. Kluwer Academic, 2000.

[GM+94] Bill Gardner, Keith Martin, et al. Hrtf measurements of a kemar dummy-head microphone. *Massachusetts Institute of Technology*, 280(280):1–7, 1994.

[GRS+10] Ian Gibson, David W Rosen, Brent Stucker, et al. *Additive manufacturing technologies*. Springer, 2010.

[HAT] Hats produced by brüel & kjær: `http://www.bksv.com/`.

[HVH+14] Tomi Huttunen, Antti Vanne, Stine Harder, Rasmus Reinhold Paulsen, Sam King, Lee Perry-Smith, and Leo Kärkkäinen. Rapid generation of personalized hrtfs. In *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*. Audio Engineering Society, 2014.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[Jen95] Rick L Jenison. A spherical basis function neural network for pole-zero modeling of head-related transfer functions. In *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*, pages 92–95. IEEE, 1995.

[JJ13] Finn Jacobsen and Peter Moller Juhl. *Fundamentals of General Linear Acoustics*. John Wiley & Sons, 2013.

[Jol02] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[JPR+07] Finn Jacobsen, Torben Poulsen, Jens Holger Rindel, Anders Christian Gade, and Mogens Ohlrich. Fundamentals of acoustics and noise control. *Ørsted: DTU, Technical University of Denmark*, 2007.

[Kat01] Brian FG Katz. Boundary element method calculation of individual head-related transfer function. i. rigid model calculation. *The Journal of the Acoustical Society of America*, 110(5):2440–2448, 2001.

[KB77] George F Kuhn and Edwin D Burnett. Acoustic pressure field alongside a manikin's head with a view towards insitu hearing-aid tests. *The Journal of the Acoustical Society of America*, 62(2):416–423, 1977.

[KEM] Kemar produced by g.r.a.s. sound and vibration: `http://www.gras.dk/`.

[Koh13] P Kohnke. Ansys mechanical apdl theory reference, 2013.

[KW92] Doris J Kistler and Frederic L Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647, 1992.

[LBJ12] Yan Luximon, Roger Ball, and Lorraine Justice. The 3d chinese head and face modeling. *Computer-Aided Design*, 44(1):40–47, 2012.

[LC87] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, volume 21, pages 163–169. ACM, 1987.

[LPC+00] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 131–144. ACM Press/Addison-Wesley Publishing Co., 2000.

[LRC03] Søren Laugesen, Karsten Bo Rasmussen, and Torben Christiansen. Design of a microphone array for headsets. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 37–40. IEEE, 2003.

[MBL07] Piotr Majdak, Peter Balazs, and Bernhard Laback. Multiple exponential sweep method for fast measurement of head-related transfer functions. *Journal of the Audio Engineering Society*, 55(7/8):623–637, 2007.

[MLGM89] JP Moss, AD Linney, SR Grindrod, and CA Mosse. A laser scanning system for the measurement of facial surface morphology. *Optics and Lasers in Engineering*, 10(3):179–190, 1989.

[Mor86] Philip McCord Morse. *Theoretical acoustics*. Princeton University Press, 1986.

[MSHJ95] Henrik Møller, Michael Friis Sørensen, Dorte Hammershøi, and Clemen Boje Jensen. Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5):300–321, 1995.

[MVCL12] Ryan W McCreery, Rebecca A Venediktov, Jaumeiko J Coleman, and Hillary M Leech. An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *American journal of audiology*, 21(2):295–312, 2012.

[NITI07] Takanori Nishino, Naoya Inoue, Kazuya Takeda, and Fumitada Itakura. Estimation of hrtfs on the horizontal plane using physical features. *Applied Acoustics*, 68(8):897–908, 2007.

[NUR] Geomagic studio: `www.geomagic.com`.

[OHI09] Makoto Otani, Tatsuya Hirahara, and Shiro Ise. Numerical study on source-distance dependency of head-related transfer functions. *The Journal of the Acoustical Society of America*, 125(5):3253–3261, 2009.

[PBL10] Rasmus R Paulsen, Jakob Andreas Bærentzen, and Rasmus Larsen. Markov random field surface reconstruction. *Visualization and Computer Graphics, IEEE Transactions on*, 16(4):636–646, 2010.

[PEK+12] OB Pedersen, Christian Erikstrup, Sebastian Ranzi Kotze, E Sø rensen, MS Petersen, Katrine Grau, and Henrik Ullum. The danish blood donor study: a large, prospective cohort and biobank for medical research. *Vox sanguinis*, 102(3):271–271, 2012.

[PKA+09] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference On*, pages 296–301. IEEE, 2009.

[PL10] Rasmus R Paulsen and Rasmus Larsen. Anatomically plausible surface alignment and reconstruction. In *Theory and Practice of Computer Graphics*, pages 249–254, 2010.

[PLN+02] Rasmus Paulsen, Rasmus Larsen, Claus Nielsen, Sø ren Laugesen, and Bjarne Ersbø ll. Building and testing a statistical shape model of the human ear canal. In *Medical Image Computing*

and Computer-Assisted Intervention-MICCAI 2002, pages 373–380. Springer, 2002.

[Ras07] Karsten Bo Rasmussen. Microphone system with directional response, May 1 2007. US Patent 7,212,642.

[Ray07] Lord Rayleigh. Xii. on our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232, 1907.

[RBW95] PR Runkle, MA Blommer, and GH Wakefield. A comparison of head related transfer function interpolation methods. In *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*, pages 88–91. IEEE, 1995.

[RC11] Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011.

[RM99] Todd Ricketts and H Gustav Mueller. Making sense of directional microphone hearing aids. *American Journal of Audiology*, 8(2):117–127, 1999.

[RMBB08] Radu Bogdan Rusu, Zoltan Csaba Marton, Nico Blodow, and Michael Beetz. Learning informative point classes for the acquisition of object model maps. In *Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on*, pages 643–650. IEEE, 2008.

[RV05] S. Romdhani and T. Vetter. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 986–993. IEEE, 2005.

[SHB08] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Thomson, 2008.

[SR93] RW Stadler and WM Rabinowitz. On the potential of fixed arrays for hearing aids. *The Journal of the Acoustical Society of America*, 94(3):1332–1342, 1993.

[SST06] Rod R. Seeley, Trent D. Stephens, and Philip Tate. *Anatomy & physiology*. McGraw-Hill, 2006.

[str] Stratasys polyjet objet500 connex 3d printer: `www.stratasys.com`.

[TL94] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 311–318. ACM, 1994.

[VBS02] Barnabas Victor, Kevin Bowyer, and Sudeep Sarkar. An evaluation of face and ear biometrics. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 429–432. IEEE, 2002.

[VFP95] Michael Valente, David Fabry, and Lisa G Potts. Recognition of speech in noise with hearing aids using dual microphones. *Journal of the American Academy of Audiology*, 6(6), 1995.

[VRSV11] RTA Van Rootseler, LJ Spreeuwers, and RNJ Veldhuis. Application of 3d morphable models to faces in video images. 2011.

[WGS11] Hagen Wierstorf, Matthias Geier, and Sascha Spors. A free database of head related impulse response measurements in the horizontal plane with multiple distances. In *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.

[WK89] Frederic L Wightman and Doris J Kistler. Headphone simulation of free-field listening. i: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2):858–867, 1989.

[WK97] Frederic L Wightman and Doris J Kistler. Factors affecting the relative salience of sound localization cues. *Binaural and spatial hearing in real and virtual environments*, 1:1–23, 1997.

[YB07] Ping Yan and Kevin W Bowyer. Biometric recognition using 3d ear shape. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(8):1297–1308, 2007.

[ZDGG06] Dmitry N Zotkin, Ramani Duraiswami, Elena Grassi, and Nail A Gumerov. Fast head-related transfer function measurement via reciprocity. *The Journal of the Acoustical Society of America*, 120(4):2202–2215, 2006.

[ZHDD03] Dmitry N Zotkin, Jane Hwang, R Duraiswaini, and Larry S Davis. Hrtf personalization using anthropometric measurements. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 157–160. Ieee, 2003.