Evaluation of 3D Positioned Sound in Multimodal Scenarios

Møller, Anders Kalsgaard

DOI (link to publication from Publisher):
10.5278/vbn.phd.engsci.00146

Publication date:
2016

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

EVALUATION OF 3D POSITIONED SOUND IN MULTIMODAL SCENARIOS

BY
ANDERS KALSGAARD MØLLER

DISSERTATION SUBMITTED 2015

AALBORG UNIVERSITY
DENMARK

Evaluation of 3D Positioned Sound in Multimodal Scenarios

Ph.D. Dissertation

By Anders Kalsgaard Møller

Dissertation submitted August 8, 2016

Thesis submitted:	August 8, 2016
PhD Supervisor:	Prof. Dorte Hammershøi Aalborg University
Co-PhD Supervisor:	Assoc. Prof. Flemming Christensen Aalborg University
PhD Committee:	Prof. Wolfgang Ellermeier, Technische Universität Darmstadt Senior Lecturer. Arne Nykänen, Luleå University Associate Prof. Christian Sejr Pedersen (Chairman), Aalborg University
PhD Series:	Faculty of Engineering and Science, Aalborg University

ISSN (online) - 2246-1248
ISBN (online) - 978-87-7112-399-9

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2016

Biography

Anders Kalsgaard Møller

Anders Kalsgaard Møller is a Ph.D. student at Aalborg University and has in his Ph.D. fellowship been enrolled at the department of electronic systems. Anders was born in 1987 in a small city named Mejrup in the western part of Jutland in Denmark. Anders went to high school in Lemvig. He moved to Aalborg to take a bachelor and later a master degree in Engineering Psychology at Aalborg University. In his bachelor project he worked with synthetic speech and a communication device for disabled people. In his master thesis he worked with categorization of menu structures in graphical user interfaces and how the location of an object could affect the usability.

Abstract

Our hearing enables us to locate sound sources, communicate and hear auditory warnings, which plays a significant role in our daily lives. Using binaural techniques the sound pressure at our two ears could be recreated and thereby create or recreate a 3D sound experience. This could be used in situations, where the natural auditory informations are limited but where the need to locate objects and communicate still exists.

The objective of the thesis was to identify ways of implementing 3D sound in selected multimodal applications, and evaluate the feasibility of using 3D sound in these applications. The applications are: 1) A teleconference application and 2) A 3D sound system in trucks.

The thesis consists of seven papers. Paper A-D deals with experiments with a teleconference application in which one participant is not physically present (the visitor), but interacts with the other meeting participants using different virtual reality technologies, such as stereoscopic video and binaural 3D sound. Paper A-C describes work on a solution called a hear-through device, which is intended to make it possible to play the sounds from the visitor through earphones without attenuating the sound from the other participants. This is done by placing a microphone on the outside of the earphones, that record the sound from the environment and plays it back through the earphones. The results of Paper A and B showed that the auditory information is preserved for frequencies lower than 4-5kHz. In Paper B and C, it was shown that this does not affect the quality of communication and speech-on-speech spatial release from masking, but affects localization ability and the naturalness of the sounds.

In Paper D it was evaluated if the visitor's experience of a teleconference application was affected by the type and quality of the auditory and visual inputs. In the specific situation, with only two other persons sitting in front of the listener, 3D sound did not improve the communication.

In Paper E-G it was investigated if 3D-sound could be used to give the truck driver an audible and lifelike experience of the cyclists' position, in relation to the truck, so the truck driver had an intuitive awareness of the cyclist's position. Thus, it was attempted to raise the truck driver's situational awareness so the driver could make the right decision. The 3D sound gave the truck drivers a sense of security, in particular in cases in which the cyclists came from behind, and the driver hears the cyclists, before he/she spots them. In these situations, the truck drivers reacted as intended by checking the mirror, and acting accordingly.

The conclusion on the objective is that the implementation of the 3D sound in the truck improved the situational awareness of the truck driver, and can therefore be deemed feasible for the application. Implementations of 3D sound in teleconference application can have an effect as long as the task is sufficient challenging.

Resumé

Vores evne til at lokalisere lydkilder med hørelsen, kommunikere og høre advarsler spiller en væsentlig rolle i vores dagligdag. Ved hjælp af binaurale metoder kan man genskabe lydtrykkene ved ørerne og dermed skabe eller genskabe en 3D lydoplevelse. Dette kan være brugbart i situationer, hvor den naturligt fremkommende auditive information er begrænset, men hvor der stadig kan være brug for at lokalisere objekter eller kommunikere med andre.

Formålet med afhandlingen er at afdække anvendelsesmulighederne for 3D lyd i udvalgte multimodale applikationer og samtidig evaluere 3D-lydens indvirkning i disse applikationer. De udvalgte applikationer er: 1) En telekonference applikation og 2) Et 3D lydsystem i lastbiler.

Afhandlingen består af syv artikler. Artikel A-D omhandler eksperimenter med et telekonference møde, hvor en deltager ikke fysisk er tilstede (fjerndeltageren), men interagerer med de øvrige mødedeltagere ved hjælp af forskellige virtual reality teknologier, såsom stereoskopisk video og binaural 3D-lyd. I Paper A-C arbejdes der på en løsning kaldet en hear-through device, som har til formål at gøre det muligt at afspille lyden fra fjerndeltageren over høretelefoner, uden at høretelefonerne skærmer for lyden fra de øvrige deltagere. Dette gøres ved at placere en mikrofon uden på hovedtelefonerne, som optager lydene fra omgivelserne og afspiller dem igen gennem hovedtelefonerne. Resultaterne fra artikel A og B viste, at den akustiske information er upåvirket for frekvenser under 4-5kHz. I artikel B og C blev det vist, at dette ikke påvirker kvaliteten af kommunikationen og opfattelsen af talegenkendeligheden, men påvirker lokaliseringsevnen og klangfarven.

I artikel D arbejdes der med telekonference mødet fra fjerndeltagerens side, hvor det undersøges, hvordan fjerndeltagerens oplevelse påvirkes af forskellige typer af video- og lydgengivelser. I testen med kun to andre mødedeltagere siddende foran lytteren, vurderes det, at 3D-lyd ikke forbedrede kommunikationen.

I artikel E-G anvendes 3D-lyd til at give lastbilchaufførerer en naturtro lydoplevelse af cyklisterne nær lastbilen, for at højne chaufførens fornemmelse af cyklerne, så chaufføren kan træffe de rigtige beslutninger i forbindelse med f.eks. højresving. 3D-lyden gav lastbilchaufførerne en ekstra sikkerhed især i de tilfælde, hvor en cyklist kommer bagfra, og chaufføren hører cyklisten, før han/hun ser cyklisten. I disse situationer reagerede chaufføren hensigtsmæssig og orienterede sig roligt i spejlet, fandt cyklisten og agerede herefter.

Det kan konkluderes, at brugen af 3D-lyd i lastbilen forbedrede lastbilchaufførernes fornemmelse af cyklerne. Implementeringer af 3D-lyd i telekonference systemer har en effekt så længe opgaverne er udfordrende nok.

Contents

Thesis Details

Thesis Title:	Evaluation of 3D Positioned Sound in Multimodal Scenarios
Ph.D. Student:	Anders Kalsgaard Møller
Supervisors:	Prof. Dorte Hammershøi, Aalborg University
	Assoc. Prof. Flemming Christensen, Aalborg University

The main body of this thesis consist of the following papers.

[**??**] Anders Kalsgaard Møller, Flemming Christensen, Pablo Faundez Hoffmann, Dorte Hammershøi, "Directional Characteristics for Different In-Ear Recording Points", *Proceedings of the AES 58th International Conference, 2015.*

[**??**] Anders Kalsgaard Møller, Pablo Faundez Hoffmann, Flemming Christensen, Dorte Hammershøi, "Evaluation of a Hear-through Device", *Proceedings of HCI International 17th Conference on Human-Computer Interaction, 2014.*

[**??**] Pablo F. Hoffmann, Anders Kalsgaard Møller, Flemming Christensen, Dorte Hammershøi, " Sound Localization and Speech Identification in the Frontal Median Plane with a Hear-Through Headset", *Proceedings of Forum Acusticum, 2014*,

[**??**] Anders K. Møller, Pablo F. Hoffmann, Marcello Carrozzino, Claudia Faita, Giovanni Avveduto, Franco Tecchia, Flemming Christensen, Dorte Hammershøi "Joint Evaluation of Communication Quality and User Experience in an Audio-Visual Virtual Reality Meeting", *Proceedings of Perceptual Quality of Systems, 2013.*

[**??**] Flemming Christensen, Anders K. Møller, Dorte Hammershøi "A Dynamic Binaural Synthesis System for Investigation Into Situational Awareness for Truck Drivers ", *PROCEEDINGS of the 22nd International Congress on Acoustics, 2016*.

[**??**] Anders K. Møller, Flemming Christensen, Dorte Hammershøi "Pilottest af 3D-Lydsystem i Lastbiler til Forebyggelse af Højresvingsulykker ", *Proceedings from the Annual Transport Conference at Aalborg University, 2016*.

[**??**] Anders K. Møller, Flemming Christensen, Dorte Hammershøi "A Study of a 3D Auditory Information System in Trucks to Prevent Right-Hand Turn Accidents", *Journal of Applied Ergonomics, 2016*.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty.

Preface

This thesis is submitted to the Faculty of Engineering and Science at Aalborg University. The work has been carried out in the time period September 2012 - July 2016 at Section of Acoustics (till June 2014), then at the merged section embrassing Acoustics and Signal and Information Processing, Department of Electronic Systems, Aalborg University.

The thesis starts with an introduction of the key terms and the goal of the study. A summary of the papers included are given followed by a general discussion and a conclusion wrapping up the findings.

The research carried out in this thesis is partly financed by the EU FP7 BEAMING project, project number: 248620 and by Trygfonden with the project: 3D sound in trucks (3D lyd i lastbiler). The goal of the BEAMING project was to instantaneously transport people (visitors) from one physical place in the world to another (the destination) so that they can interact with the local people there. This is achieved through shifting their means for perception into the destination, and decomposing their actions, physiological and even emotional state into a stream of data that is transferred across the internet. The goal of the Trygfonden project is to study the feasibility of using 3D sound to increase the truck driver's awareness of nearby cyclists.

I would like to thank my supervisors Dorte Hammershøi and Flemming Christensen for their guidance. Thanks to Pablo Faundez Hoffmann for his collaboration in our shared research and also thanks to the rest of the "AAU BEAMING group": Søren Krarup Olesen, Esben Madsen and Milos Markovic. Thanks to Harry Lahrmann for his role in the Trygfonden project. I wish to thank Claus Vestergaard Skipper and Peter Dissing for their aid in the lab. Thanks to Anders Tornvig Christensen for our fruitful discussions during our office time (sometimes of questionable relevans). Finally I would like to thank all the people who participated in my experiments.

Anders Kalsgaard Møller
Aalborg University, August 24, 2016

Introduction

The ability to localize sound sources, communicate and hear auditory warnings plays a significant role in our everyday tasks. When a person tries to catch our attention, when we walk in traffic or when we're talking together at a party, we use our hearing and our ability to localize sound sources.

Our sound perception is based on the sound pressures at the right and left ear (Møller 1992; Blauert 1997). Different methods have been used to record or recreate the sound pressures for example using artificial heads (Burkhard and Sachs 1975; Møller et al. 1999; Christensen, Jensen, and Møller 2000) or by recording it directly in the ears of the subject (Møller et al. 1996; Bronkhorst 1995; Blauert 1997). It has been investigated how reflections from head, pinna and shoulders affects the sound that reaches the ear. This information is known as the head related transfer functions (HRTF) (Møller et al. 1995; Blauert 1997) and are defined as: head, pinna and shoulder reflection, as compared to an omnidirectional microphone positioned in the center of the head (without the head present). The captured signals are typically played back to the listener through headphones after equalizing the signals to compensate for the headphone transfer function. The sound produced by these methods are sometimes referred to as 3D sound or binaural sound and the methods for reproducing it is referred to as binaural techniques.

One of our hearing system's primary tasks is to navigate our vision (Perrott et al. 1990). Compared to the size of our visual field the auditory field is a lot more extensive, making it possible to detect sound-producing objects outside of our visuel field. The vestibular system helps us keep track of our spatial orientation and head rotation, and the proprioceptive system help us understand how our head is turned compared to the rest of our body (Andersen et al. 1993). The interaction between our hearing, the vestibular system and proprioception enables us to quickly register the direction of a sound source and turn our vision towards it, by during a series of coordinated head and eye movements (Bolia 2004).

Our spatial hearing also allows us to segregate speech from a competing speech signal. By focusing on speech from one spatial location we can ignore sounds from background noise, and hereby improve the speech intelligibility (Bronkhorst 2000). The speech intelligibility can be further improved by moving the person speaking into our visual field, so we can use the lip movements and facial expressions to comprehend what the person is saying (Macdonald and McGurk 1978).

Our spatial hearing and the methods for reproducing the auditory information has many potential applications. This thesis includes work on two specific applications. The first application is a virtual reality project named BEAMING. The other application focus on a solution where cyclists are represented with 3D sound to truck drivers with the purpose of lowering right hand turn accidents.

BEAMING

In the EU FP7 integrated project BEAMING the goal was to instantaneously transport a person (the visitor) from one physical place in the world to another (the destination) so the visitor can interact with the people there (the locals). With different techniques, used to create a virtual reality world, peoples' senses are stimulated so they, at least to some degree, get the experience of interacting with other people as if they were physically present at the remote location (the destination).

The BEAMING project framed four different applications: A teaching application, a medical/healthcare application, a journalism and a teleconference application. The present thesis includes part of the work made in connection with the teleconferencing application.

In the teleconference application the visitor was placed in a virtual reality CAVE (Cruz-Neira et al. 1992). The CAVE consisted of three walls and a floor. A meeting room (the destination) and the meeting members were presented with a stereoscopic image on the walls and the floor of the CAVE. The audio from the destination was presented to the visitor through equalized headphones using a binaural synthesis to provide the spatial information to the visitor.

At the destination the locals could see the visitor through a tablet using an augmented reality technique. The locals were wearing an earpiece (a hear-through device) that consisted of a pair of earphones with microphones mounted on them. The sounds of the visitor were played through the earphones using a binaural synthesis to provide the cues for the perception of

direction. To avoid attenuation due to the earphone, the sounds from the other locals were captured with the microphone mounted on the earphones and played back to the locals through the earphones. The earphones and microphones were connected to a computer which was used for the recording, processing and playback. Both the locals and the visitor were wearing a headtracker that could provide the necessary tracking information about head location and rotation for the visual and auditory rendering. Figure 1 shows an example of what the visitor's side looks like and Figure 2 shows a picture of the setup at the destination, where two locals are both wearing a hear-through device.

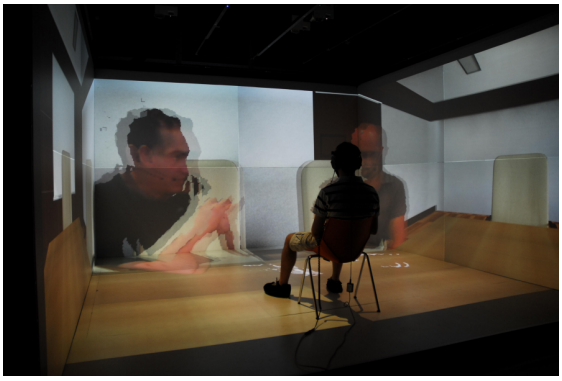

Fig. 1: The setup at the visitor's side in the CAVE. The image looks blurry due to the stereoscopic image. The visitor is wearing 3D glasses, headphones and a headtracker.

Fig. 2: The setup at the destination with the physical meeting room. The two locals were wearing hear-through devices including headpieces for the the electromagnetic tracking system. The tracking system can be seen in the foreground of the image. In the back a big screen used to present the visitor can be seen.

Previous studies of 3D sound's effect in virtual reality have indicated that 3D sound improves subjects' ability to navigate in the virtual worlds (Hendrix and Barfield 1996; Larsson, Vastfjall, and Kleiner 2002), while the effect on the experience of presence is more disputable (Hendrix and Barfield 1996; Larsson, Vastfjall, and Kleiner 2002; Larsson et al. 2007).

Similar devices to the hear-through device has previously been described as augmented reality audio (ARA) headsets (Härmä et al. 2004). In (Tikander 2005) nine subjects evaluated the sound quality of wearing an ARA-headset. In the setup the subjects were wearing an ARA-headset on the right ear and had both ears covered by circumaural headphones. Different sound samples were played through the canals of the circumaural headphones one at a time. The subjects were told to compare the sound quality between the left ear with no ARA-headset and the right ear with the ARA-headset on a Mean opinion scale [MOS) (ITU-T 2001). The sound quality was rated close to four on the MOS-scale (perceptible, but not annoying degradation).

In (Tikander et al. 2008) four subjects were given a hear-through headset in order to assess whether they would be willing to wear the headset in an everyday life situation. The subjects wore the headsets in an evaluation period that lasted between four days and one week. In that period the subject wore the headset between 20 and 35 hours. The comments were overall positive and the sound quality was very good. It was however noted that the users' own voice sounded very boomy and the sound was localized inside the head.

3D Sound in Trucks

Every year in Denmark various fatal accidents happens where a right-hand turning truck hits a cyclist driving straight ahead in an intersection. The truck driver is placed in a closed cabin with a window in front and a window to each side. The truck driver is typically positioned about 2.5 meters from the ground. To cover all the blind spots around the truck, the truck is equipped with four mirrors. If the mirrors are correctly adjusted the truck driver should have visibility to a large area along the side and in front of the truck. Despite the four mirrors the accidents still happen.

According to a technical report from the Danish Transport Agency (Trafikstyrelsen 2014) the reason for the accidents could be that the driver has to keep visual attention to cyclists, through all mirrors and windows, while paying attention to other road users, and the layout of the road. This might exceed the cognitive capability of the drivers. Figure 3 gives an impression of the demanding task it is to know if any cyclists are present near the truck. The transparent version of the cyclists with dotted lines indicates the direc-

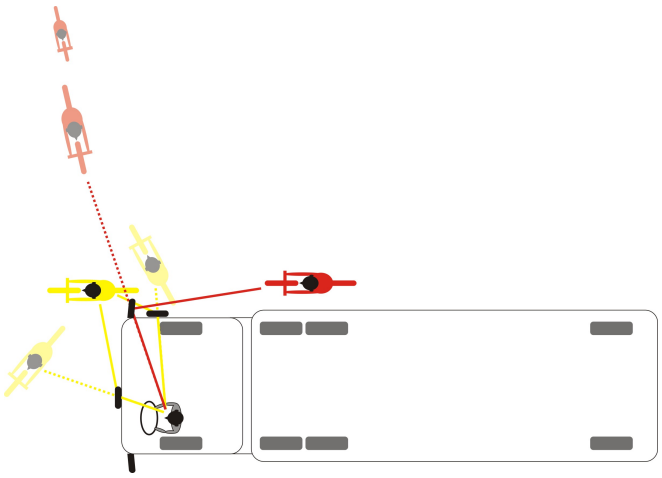

Fig. 3: Example of how the truck driver sees the cyclists. The red and yellow cyclists indicate the cyclist's actual position and the transparent version shows where the truck driver sees them

tion in which the truck driver sees the cyclists in the mirrors. The truck driver has to use cognitive resources on mapping the cyclists, the truck driver sees in the mirrors, to their actual position (indicated with the red and yellow cyclist). The real situation is even more complex because the cyclists are moving and will often only be visual in a mirror for a short time window. During the shift between the mirrors the truck driver might fail to spot the cyclist, with a possible fatal outcome as the result. By providing the truck driver with a 3D sound representation of the cyclist, the truck driver would receive continuous information about the cyclist's position in a manner that could potentially lower the cognitive load needed to keep track of the cyclist. The idea of using sound in vehicles is not new. Sound have previously been used in warning systems in cars and trucks (Fagerlönn 2010; Graham 1999). 3D sound has been used in airplanes where 3D audio was linked to visual targets to help pilots to respond faster to incoming objects such as missiles or other air crafts (Begault 1993; Veltman, A., and Bronkhorst 2004).

The aim of the solution, proposed in this thesis is, however, not to warn the truck drivers about critical situations, but make the truck driver aware of the cyclists positions, so the truck driver can make the right decision in the situation. The term situational awareness has been used by (Endslev 1996) to describe: "A person's perception of the relevant elements in an environment as determined from system display or directly by the senses". By giving the truck drivers an ecologically valid auditory representation of the cyclists, whenever they approach the truck, it is attempted to increase the truck drivers' situational awareness.

1 Objectives and Research Questions

The objective of the thesis was to identify ways of implementing 3D sound in selected multimodal applications, and evaluate the feasibility of using 3D sound in these applications. The applications are; 1) a teleconference application and 2) a 3D sound system in trucks. In connection with the objective, the following research questions were raised:

1. To what extend can the spatial information be preserved when using a hear-through device?
2. How does the hear-through affect our listening experience?
3. How important is spatial information for communication in multimodal scenarios?
4. Can 3D sound be used as an auditory display to improve truck drivers' situational awareness regarding nearby cyclists?

2 Summary of Papers

The thesis consists of seven papers. Paper A-C include research conducted in relation to the hear-through device. Research question 1 is addressed in Paper A and B and research question 2 is addressed in Paper B and C. Paper A focus on measuring the directional dependence in the sound transmission to the hear-through device with different microphone positions. The measurements are done with human subjects in a unique baffle set-up. Paper B also briefly describes the experiment from Paper A and continues with an experiment carried out to test if the hear-through device would compromise the listening experience. In paper C an evaluation of the significance of wearing a hear-through device in a localization and a speech identification task is described. Research question 3 is addressed in Paper D. Paper D describes an experiment that assess the feasibility of using 3D sound in a teleconference application by comparing it to a diotic representation of the sound. Research questions 4 is addressed in Paper E-G. Paper E describes the development of a dynamic binaural synthesis system developed for testing, if 3D sound can be used to improve truck drivers' situational awareness regarding cyclists. Paper F and G describes a field study where the system is tested. Paper G also describes a listening experiment carried out to study which auditory icons are the most suitable to be used to inform a truck driver that a cyclist is present near the truck.

3 Summary of Paper A: Directional Characteristics

This paper focus on measuring the directional dependence in the sound transmission to different microphone positions on the outside of a hear-through device. The hear-through device consists of two microphones mounted on a pair of earphones and the electronic control of recording and playback. The hear-through device enables the listener to listening to the acoustics of the surroundings, which would otherwise be attenuated by the passive attenuation of the earphones.

The aim of the study was to find the best position to place the microphone to preserve as much of the spatial information as possible and to investigate how much the directional characteristics vary between subjects. Based on measurements with an artificial ear three microphone positions were selected and measured with nine subjects. The three microphone positions and the reference position at the blocked entrance to the ear canal are shown in Figure 4. The blocked entrance to the ear canal is chosen as a reference point, because all spatial information is preserved, when recording at this point(Hammershøi and Møller 1995). The amplitude responses for each of the subjects and microphone positions are presented in Figure 5. The standard deviations across directions as a function of frequency for each of the three microphone positions are presented in Figure 6 and the mean differences from the reference point for all directions in Figure 7.

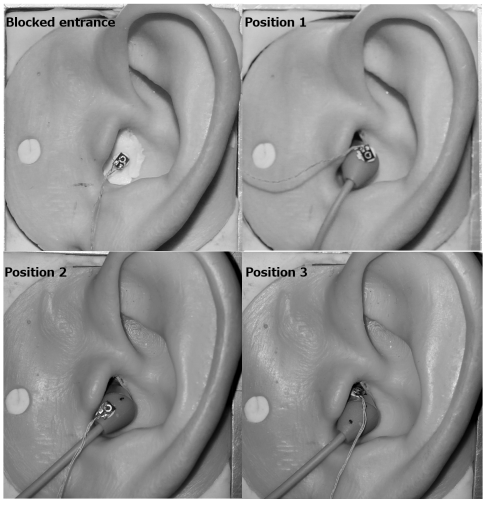

Fig. 4: The microphone positioned at the blocked entrance and at the three selected positions.

For all three microphone positions the amplitude responses for each of the 25 directions were very similar up to 4-5 kHz, where they started to deviate. Small deviations in mean could also be seen for some subjects from around 4 kHz. There was no clear difference between the three microphone positions and none of the microphone positions stood out as the best or worst choice.

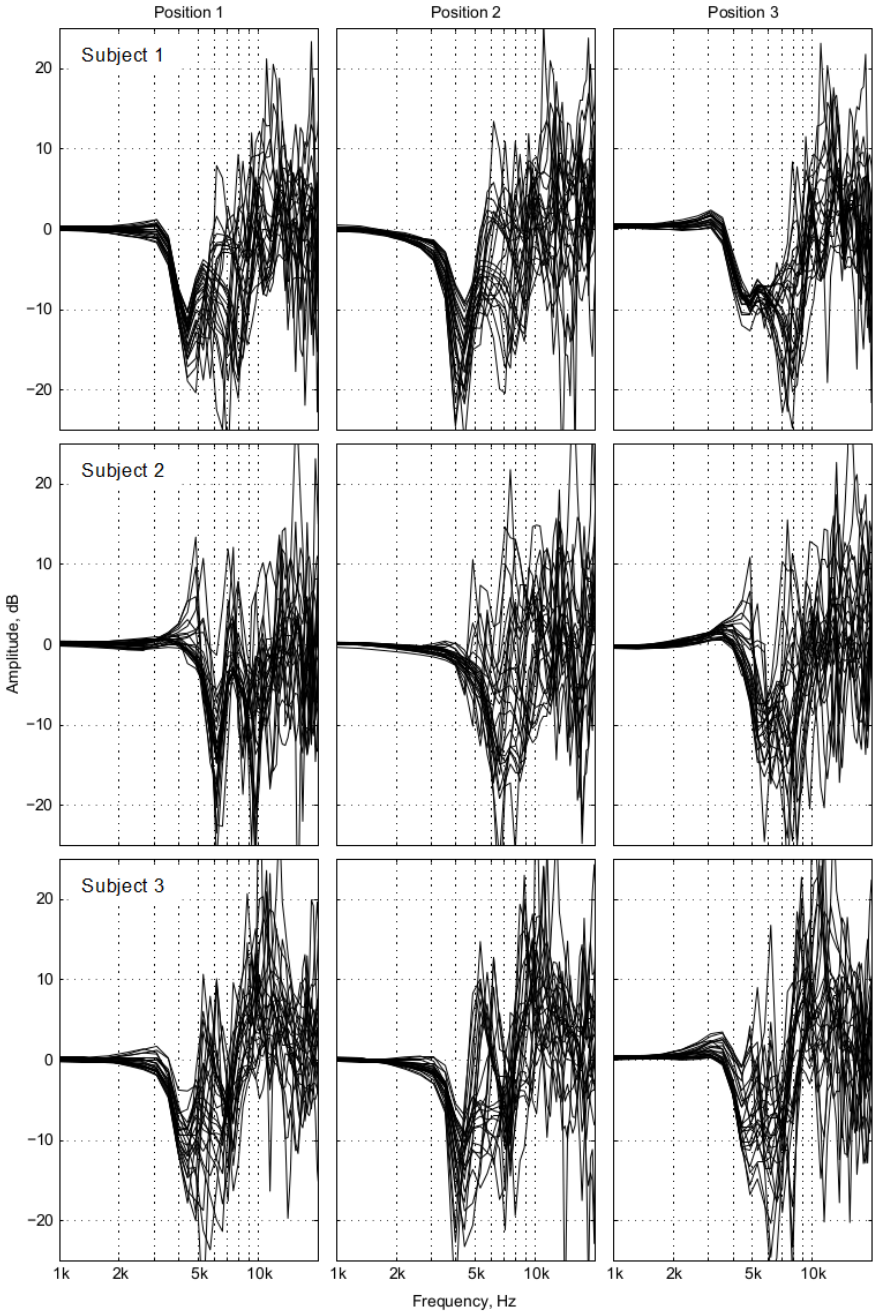

Fig. 5: Frequency response for sound transmission from the 25 different directions measured to the three different microphone positions for all nine subjects.

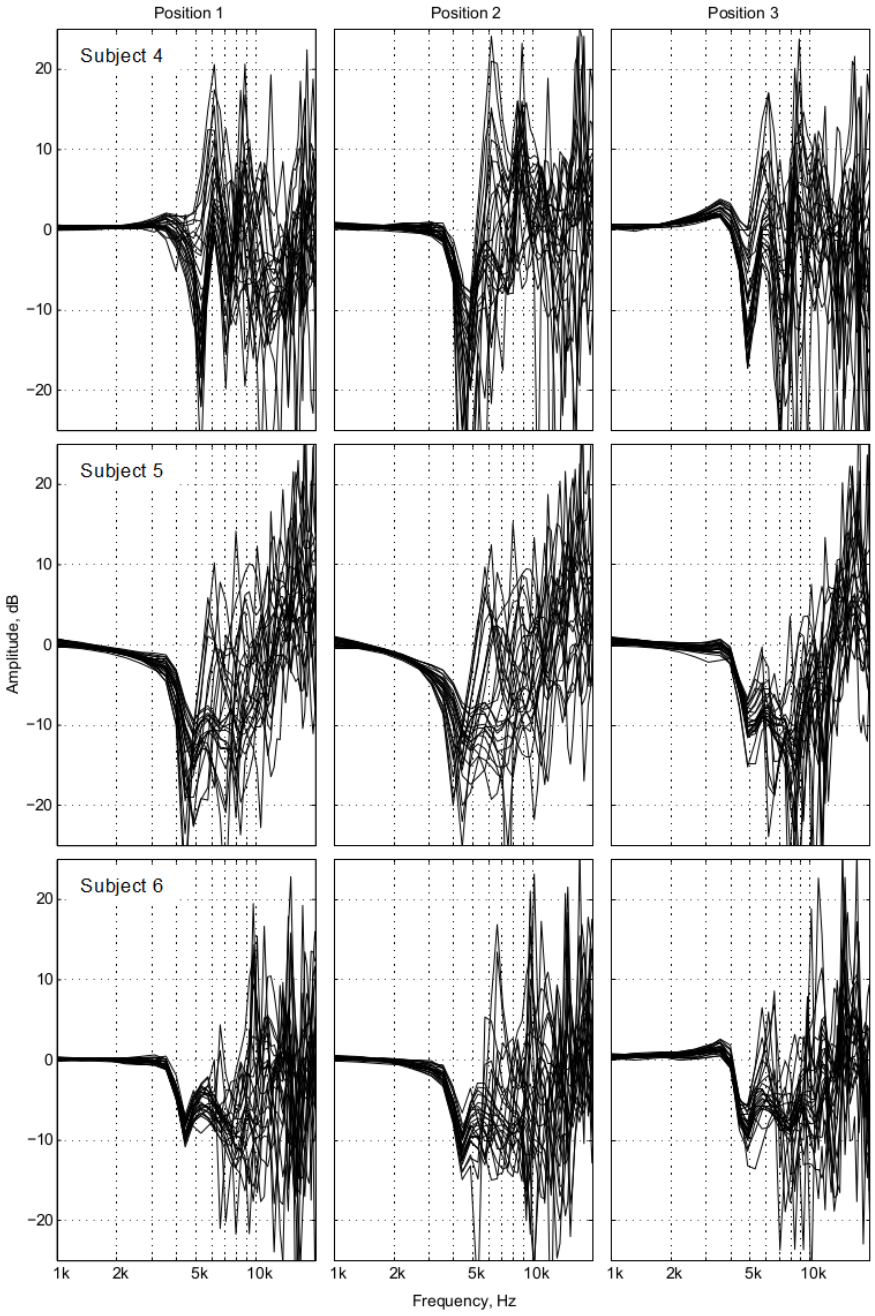

Fig 4. Continued

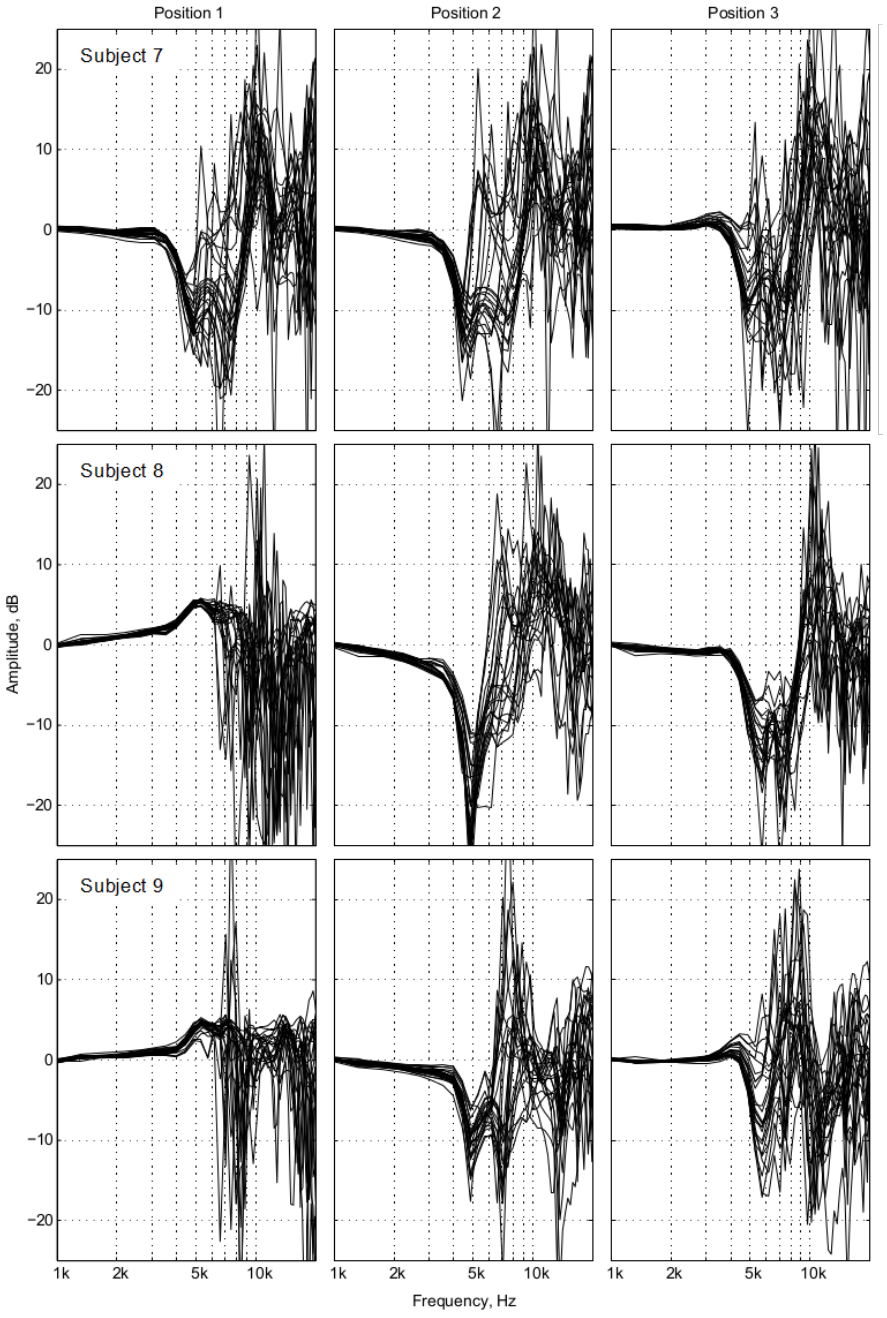

Fig 4. Continued

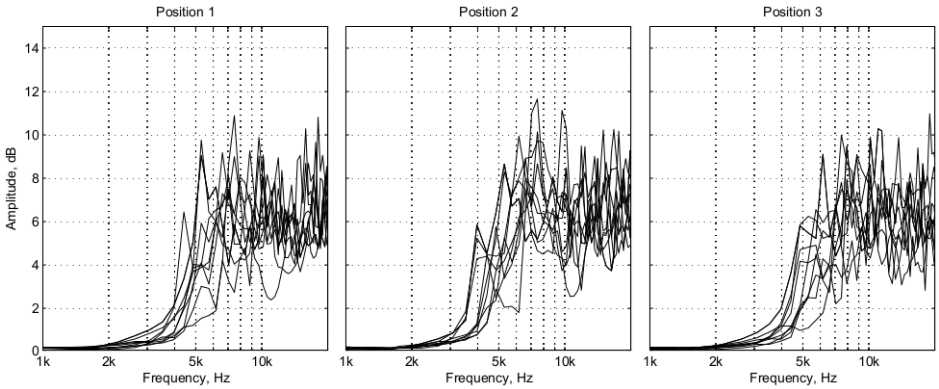

Fig. 6: Standard deviation across directions for each subject, and for all three microphone positions. Each line represents the standard deviation across 25 directions for one subject. Data for all nine subjects.

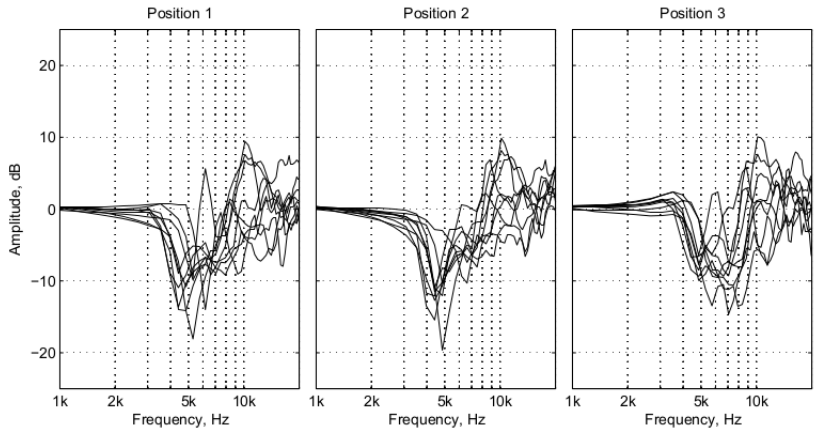

Fig. 7: Mean difference from the reference point (the blocked entrance) for the 25 directions for each subject, and for all three microphone positions. Each line represents the mean difference from the reference point for 25 directions for one subject. Data for all nine subjects.

4 Summary of Paper B: Evaluation of a Hear-Through

The experiment presented in Paper A is the same as in Paper B. The results in Paper B is, however, reported at a preliminary stage when only five subjects had been measured. By mistake the wrong type of earphones was mentioned in this paper. The correct type is AKG 323 xs Blue as it is described in Paper A.

The focus in Paper B is the experiment, where a prototype of the hear-through device was used to study, whether the deviations in the reproduced sound affect the listener experience. The evaluation involved groups consisting of three subjects, who were instructed to solve a puzzle in a joint effort. The task was chosen, because it encourages the subjects to move around and to communicate with each other. Each group is given the task three times, once wearing the hear-through device (hear-through on), once with the natural condition (control), and once with an occluded condition (with the ears blocked with earphones and the hear-through turned off). After completing each session, the subjects were given a questionnaire with six questions. The six questions were stated on Visual Analog Scales (VAS) with bipolar end-labels that were used to evaluate the listening experience. The six questions were addressing: The ability to understand what the other subjects said, the sound of the subject's own voice (if their own voice sounded natural), localization of the voice of the other subjects, the noise level, the naturalness of the sound from the surroundings, and the overall sound quality.

A pilot study was carried out with three subjects but by mistake, a flat frequency response was used as the target function for the earphone equalization. The target function for the earphone equalization should, according to (Hoffmann, Christensen, and Hammershøi 2013b) (and in line with (Møller 1992), equal the blocked entrance to 'eardrum' transfer of the coupler, when the earphone is calibrated in the coupler. The target function is presented in (Hammershøi and Møller 2008) (Fig. 2, left panel). Due to the equalization error the experiment was discontinued, and data only exists for the pilot study with the flawed equalization.

The results from the pilot test is presented in Figure 8. The figure shows how each subject rated the session based on the six questions. Each dot on the figure represents responses from one subject. The pilot study indicated that the device was not hindering for the communication or for localizing the voices of the other persons. The subjects rated the noise level as higher in the hear-through session. This was not unexpected, as the miniature microphones have high ambient noise levels. The ratings also indicated that

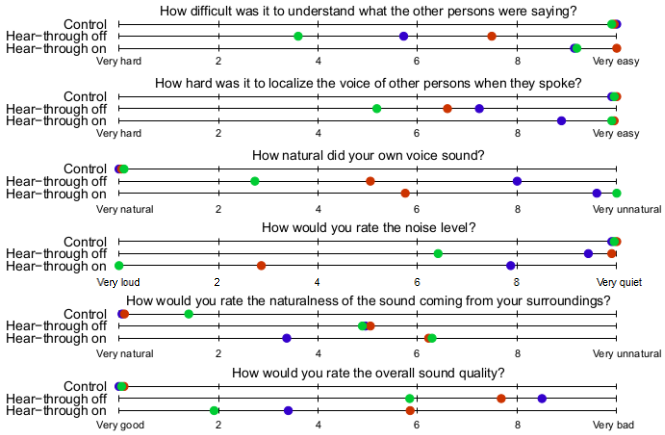

Fig. 8: The results from the pilot study. The subjects rated each session on all six rating scales. Each dot represents responses from one subject.

the sounds were unnatural in the hear-through condition, especially the experience of the subjects' own voices. These issues seem to have affected the overall sound experience, which was rated lower than the control session, but not as low as the occluded condition.

In the interview after the experiment, the subjects reported that their own voice sounded "boomy" and that some sounds like shoes rubbing against the carpet sounded unnatural. This is not a surprising result. The occlusion effect is well known from the hearing aids users and can be solved by one of the following methods; 1) adding a small leak ("vent") that allows the low frequencies in the ear canal to equalize with the surroundings, and 2) an active attenuation of the unwanted signal in the ear canal, which comes from the bone structure. The first solution is simple and will certainly be applicable in many scenarios.

The error that was introduced in the headphone equalization, resulted in a mismatch between the desired frequency response, and the actual frequency response. This deviation was the same for all directions, and would thus be expected result in a discoloration of the sound, thus affecting question five. Humans do, however, adapt successfully to static discolorations in e.g. radio and phone transmissions, and it is presumed that the discoloration had probably only minor influence in the present, spatially dynamic case. The hear-through system thus demonstrates a feasible way forward for mixed reality scenarios.

5 Summary of Paper C: Sound Localization and Speech Identification

An important aspect of the hear-through headset is how close to real life the sounds are perceived. As presented in Paper A and in (Hoffmann, Christensen, and Hammershøi 2013b) wearing a hear-through device will introduce errors in the spatial information of the sound reproduced. In Paper C it is investigated how our sound perception is affected by these deviations. Two experiments were conducted in this study: A frontal vertical-plane sound localization test, and a speech-on-speech spatial release from masking test. Ten subjects (two females and eight males) participated in the experiments. The experiments were carried out in an anechoic room. The setup consisted of seven loudspeakers, five placed at 0° azimuth in front of the subjects at 0°, $\pm$ 22.5°, and $\pm$ 45° elevation. The last two loudspeakers were placed at $\pm$ 45° azimuth and 0° elevation.

In the sound localization experiment a 500-ms white-noise burst was played from one of the loudspeakers, and the subjects should indicate from which loudspeaker they heard the sound. The test was conducted with three conditions; 1) a natural listening condition, 2) a hear-through condition with the hear-through on, and 3) an occluded condition with the hear-through off.

In the speech-on-speech spatial release from masking test two sentences were played simultaneously from two loudspeakers in each trial. The sentences all had the form: "ready <CALLSIGN> go to <COLOR> <DIGIT> now". The sentences came from The Air Force Research Laboratory's publicly available coordinate response measure speech corpus (Bolia et al. 2000), and consisted of four colors and seven digits. The subjects' task in the experiment were to recognize the digit and color of the sentence, where the callsign "baron" was mentioned. Each subject completed 90 trials in a natural listening condition and 90 trials in a hear-through condition.

The error rate in the frontal median plane of the sound localization experiment is presented in Figure 9. In the localization experiment the subjects performed better with the hear-trough on than with the hear-through off. The localization performance was, however, impaired by the hear-through headset relative to the natural condition. The results from the speech identification task is presented in Figure 10.

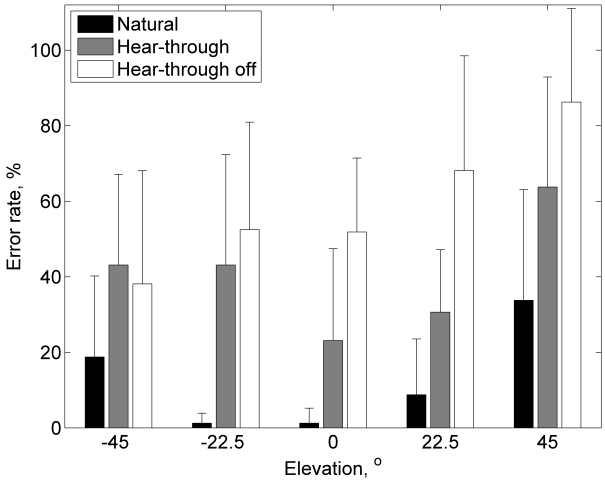

Fig. 9: The mean error rate for each loudspeaker position in the frontal median plane for each condition (n=10). Error bars indicate the standard deviation

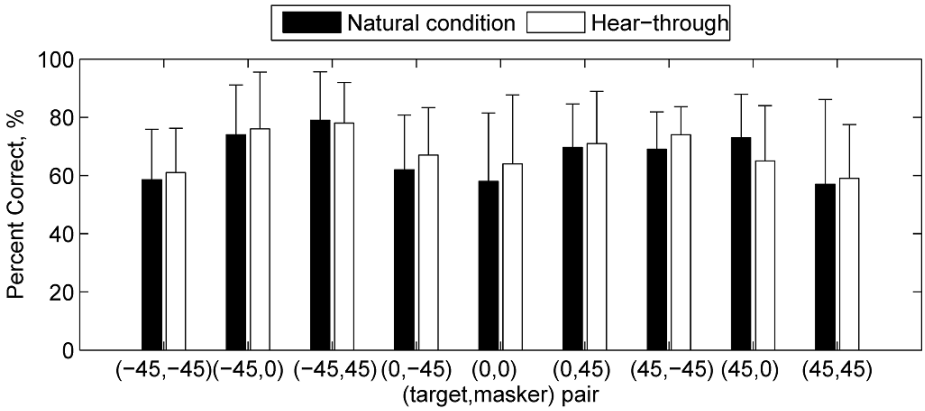

Fig. 10: The mean percentage of correct answers (n=10). Error bars indicate the standard deviation.

The results suggest that normal speech-on-speech spatial release from masking and the localization in the frontal median plane was unaffected by the use of the hear-through headset. Frontal vertical localization was, however, affected by the hear-through device. The experiments thereby confirmed that the performance of the hear-through device is acceptable for sound with energy below 4-5 kHz, where spatial information is preserved. The experiments also confirmed that critical tests including and relying on sound energy above 4-5 kHz, revealed the effect of the deteriorated spatial information for frequencies above 4-5 kHz.

With the current experimental design, where only the frontal median plane was tested, it was not possible to make any conclusions on how the hear-through device would affect the perception of sound sources placed on the same interaural time difference contour - the so-called cone of confusion (Blauert 1997). On the cone of confusion, no interaural time differences exists, therefore the listener has to rely on pinna cues and head movements to distinguish between two points on the cone. With the hear-through blocking the pinna it is very likely that this will affect the number of cone of confusion errors.

6 Summary of Paper D: Evaluation of a Virtual Reality Meeting

In this study, it was evaluated how the experience of a teleconference application is affected by the type and quality of the auditory and visual inputs. In the teleconference application the subjects were observing a meeting between two business partners planning a virtual reality experiment. Twenty-four subjects participated in the experiment, which was carried out in a virtual reality CAVE. The subjects were all Italian speaking and the whole experiment were carried out in Italian. Two different sound presentations were used (diotic and binaural), and two different video presentations using two levels of video quality. The mix between these variables created four different sessions.

During the sessions, there was a problem with an asynchrony between audio and video. The problem occurred, because the video and the audio were presented using two different software programs, and sometimes a missing data package delayed the video stream creating asynchrony between video and audio.

After each session, the subjects were given a multiple-choice questionnaire. The questionnaire contained 10 questions related to the content of the meet-

ing. After completing the questionnaire, subjects rated the session on five scales used to measure the experience. The five scales were measuring; 1) presence, 2) communication quality, 3) tiredness, 4) video quality and 5) audio quality. After the last session, subjects were debriefed.

There were very few errors in the answers to the questions regarding the content of the meeting, which indicated that it was easy, to follow the discussion between two people. The few errors could have occurred by chance, and provided a poor basis for further analysis.

Boxplots for each of the ratings are presented in Figure 11. As can be seen on the boxplots they contain a high degree of dispersion and contained a considerable amount of outliers. This is most likely due to the synchronization error between the video and the audio, which would sometimes occur in the end of the session.

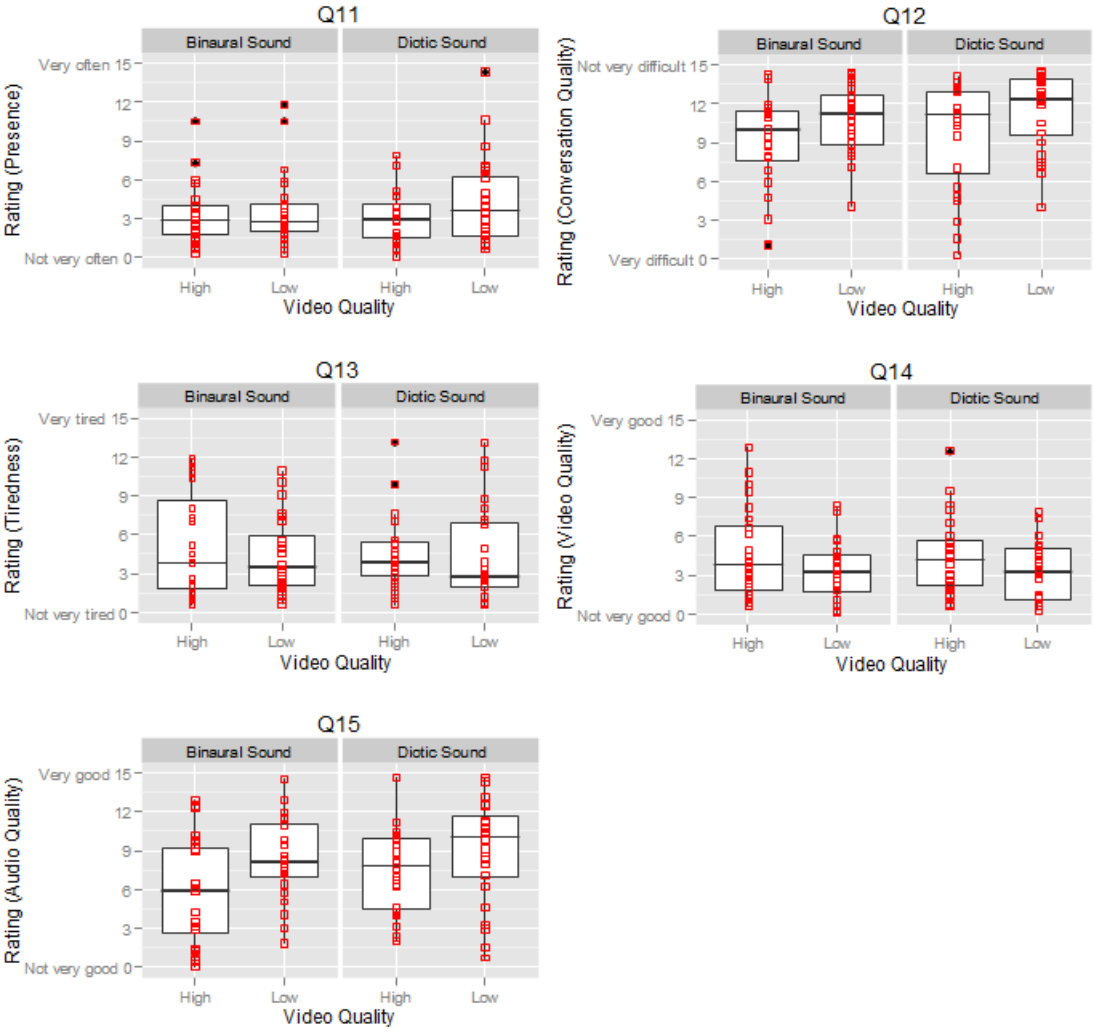

Fig. 11: Boxplot for each of the five scales used to measure the experience of each session, answers from 24 subjects. The x-axis indicates the video quality (High/low) for the Binaural and the Diotic sound representation. The median is indicated with the black line, the box showing the middle 50% of observations (interquartile) and the whiskers indicate the last observation within the 1.5 times range of the interquartile. The black dots represent the outliers (observations outside of the whiskers).

Paper D only provided some preliminary results without the proper data analysis, but this is carried afterwards and reported in the summary.

A one-way repeated measure ANOVA test was carried out to test for differences between groups for each question. Before the ANOVA test a Mauchly's test (Anderson 2003) was carried out to test for the assumption of sphericity. If the assumption was not met the degrees of freedom and thereby the p-value was adjusted using the Hyunh-Feldt estimate of sphericity (Huynh and Feldt 1976). As a post-hoc test a paired t-test was used, if the assumptions of normal distributed residuals and homogeneity of variance was met, otherwise a Wilcoxon Signed Rank test (Wilcoxon 1945) was used.

Q11:During the experience, how often did you think of yourself as actually participating in the meeting
Mauchly's test indicated that the assumption of sphericity had been violated $x^2(5)=15.26$, p=0.0093, therefore degrees of freedom was corrected using Hyunh-Feldt estimates of sphericity (ϵ=0.750). The results showed that the rating was not significantly affected by the session. F(2.25,51.7)=1.23, ω^2=0.0247, p=0.302.

Q12:How difficult was it to follow the conversation during this part of the meeting?
Mauchly's test indicated that the assumption of sphericity had been met $x^2(5)=6.00$, p=0.306. The results of the ANOVA showed that the rating was significantly affected by the session. F(3,69)=3.60, ω^2=0.0708, p=0.0176.

A Shapiro test (Rovstorn 1982) was used to test, if the residuals in all groups come from the same normal distribution, before deciding on a post hoc test. The test showed that the residuals are significantly different from a normal distribution, W = 0.9233, p-value = 3.029e-05, therefore the non-parametric Wilcoxon Signed Rank test was applied as a post hoc test. The test was applied as a pairwise comparison with a Bonferroni correction (Wright 1992) to the p-value. With a p-value = 0.045, the test showed a significant difference between the session with Low quality video and binaural sound, and the session with high quality video and binaural sound. All the p-values are show in Table 1.

Q13:To what degree did you feel tired during this part of the meeting?
Mauchly's test indicated that the assumption of sphericity had been met $x^2(5)=5.23$, p=0.389. The results of the ANOVA showed that the rating was not significantly affected by the session. F(3,69)=0.768, ω^2=0.0110, p=0.515.

	High quality video & Binaural sound	High quality video & Diotic sound	Low quality video & Binaural sound
High quality video & Diotic sound	p-value = 1.000	-	-
Low quality video & Binaural sound	p-value = 0.045	pvalue = 0.356	-
Low quality video & Diotic sound	p-value = 0.392	p-value = 1.000	p-value = 1.000

Table 1: A table of all the p-values from the pairwise comparison of the sessions using a Wilcoxon Signed Rank test with a Bonferroni correction for Q12.

Q14:How do you rate the quality of the video in this part of the meeting?
Mauchly's test indicated that the assumption of sphericity had been met $x^2(5)=8.38$, p=0.137. The results of the ANOVA showed that the rating was significantly affected by the session. F(3,69)=3.06, ω^2=0.0546, p=0.0339.

The Shapiro test showed that the residuals are significantly different from a normal distribution, W = 0.9471, p-value = 0.00072, therefore a pairwise Wilcoxon Signed Rank test with a Bonferroni correction was applied, as a post hoc test. The test showed no differences between the groups. All the p-values from the Wilcoxon test are shown in Table 2.

Q15:How do you rate the quality of the audio in this part of the meeting?
Mauchly's test indicated that the assumption of sphericity had been met $x^2(5)=6.41$, p=0.269. The results of the ANOVA showed that the rating was significantly affected by the session. F(3,69)=5.63, ω^2=0.103, p=0.00163.

A Shapiro test was used to test if the residuals in all groups come from the same normal distribution. The test showed that the residuals are not significantly different from a normal distribution, W = 0.985, p-value = 0.3463.

A Levene test (Levene 1960) was used to test for homogeneity of variance to decide on which t-test should be used (Student's t-test or Welch t-test). The test showed that the variances were not significantly different in the groups F(3,92)=0.9945, p=0.399. A pairwise paired Student's t-test with a Bonferroni adjustment was therefore used to test for differences between the groups. The test showed a significant difference between the session with Low quality video and binaural sound, and the session with high quality video and binaural sound. All the p-values from the t-test test are shown in Table 3

The subjects rated the communication quality as being better in the ses-

	High quality video & Binaural sound	**High quality video & Diotic sound**	**Low quality video & Binaural sound**
High quality video & Diotic sound	p-value = 1.00	-	-
Low quality video & Binaural sound	p-value =0.25	p-value = 0.31	-
Low quality video & Diotic sound	p-value = 0.64	p-value = 0.64	p-value = 1.00

Table 2: A table of all the p-values from the pairwise comparison of the sessions using a Wilcoxon Signed Rank test with a Bonferroni correction for Q14.

	High quality video & Binaural sound	**High quality video & Diotic sound**	**Low quality video & Binaural sound**
High quality video & Diotic sound	p-value = 0.1285	-	-
Low quality video & Binaural sound	p-value = 0.0045	p-value = 0.3646	-
Low quality video & Diotic sound	p-value = 0.0937	p-value = 1.0000	p-value = 1.0000

Table 3: A table of all the p-values from the pairwise comparison of the sessions using a Student's t-test with a Bonferroni correction for Q15.

sion with binaural sound and high quality video compared to the session with binaural sound and low quality video. The higher video quality might have made it easier for the subjects to notice the facial expressions and lip movements of the two business partners, and the communication quality was therefore increased.

The subjects rated the quality of the audio higher in the session with binaural sound and low video quality compared to the session with binaural sound and high video quality. This seems a bit odd but a potential explanation would be that the asynchrony between audio and video would cause the audio to be out of sync with the lip movements, which would be easier to spot with the higher video quality.

In the debriefing interview the subjects were asked to describe how they experienced the audio and video in each session. The comments about the video quality were mainly related to the overall image quality. A few subjects also pointed out issues with the recording of the actors like: ”a small part of the actor's body disappeared” or ”the actors appeared too big”. Fewer comments were given in relation to the audio and these were mainly addressing the synchronization error. Nobody mentioned the spatial properties of the sound, which could mean that they never realized that it changed between

the sessions. This is likely due to the fact that the two business partners as well as the subject is all seated meaning the sound sources will all be in relative fixed positions and thereby there is no need to locate the sound sources making the spatial hearing indifferent for this task.

7 Summary of Paper E: A dynamic binaural synthesis system

Paper E describes the development of a dynamic binaural synthesis system. The system was developed with the purpose of investigating if an audible 3D representation of cyclists can increase truck drivers' situational awareness. It is assumed that an increased situational awareness of the truck drivers may reduce the number of collisions in intersections between right-turning trucks and cyclists going straight ahead.

In the test system, a facilitator had to spot the cyclists and use a graphical interface to move the apparent sound position along with the cyclist's actual position. The test system should be easy to deploy in different trucks, and a solution using headphones was therefore chosen. The headphones (Sennheiser PC 363D) had an acoustically very open design so that the driver would still experience the surroundings in a close to natural way.

The hear-through earphone system was also considered for this use, as these would allow an even more natural situation for the drivers. Its imperfections with respect to background noise, and the fact that the tracking systems require a head-band for mounting, made the acoustically transparent earphones the most logical compromise for the field study.

The sound presented over the headphones is generated using a 3D sound software. The truck driver's head movements were tracked using an orientation tracker mounted on top of the headphones. Another orientation tracker was placed in the truck cabin to track the orientation of the truck. The data from the orientation trackers were used to calculate the distance and orientation of the cyclist provided by the graphical user interface (GUI). All the tracking information was used by the 3D software engine to process and generate the 3D audio.

The system consisted of four software modules. The different software modules communicated through the network protocol UDP. The head tracking module works as the server and the truck tracking, the experimenter inter-

face, and the 3D sound engine are all software modules connected as clients to the server module.

The system was running on a laptop computer and a Windows tablet connected to the laptop computer via a direct Ethernet connection. The tablet was running the experimenter interface. The rest of the system was running on – or connected to – the laptop.

Paper E also briefly describes the field study for which the system was created. This will be explained in more details in Paper F.

The test system has some limitations compared to a potential final implementation. To make a full automatic implementation of such a system, sensors would have to scan the surroundings of the truck for nearby cyclists. The sound would most likely be presented through loudspeakers.

8 Summary of Paper F: Pilottest af 3D-lydsystem i lastbiler

The aim of the study was to use 3D-sound to give the driver an audible and lifelike experience of the cyclists' position in relation to the truck, so the truck driver had an intuitive awareness of the cyclist's positions. Thus, was it attempted to raise the driver's "situational awareness". With a greater awareness of where and when there are cyclists near the truck, the driver would probably be able to use the mirrors more efficiently, and potentially avoid collision when turning right.

The system was tested with four different truck drivers on their goods delivery routes. Two of the truck drivers drove two trials with the system. The trucks were approximately 12 meters long and 2.55 meters wide.

From the passenger seat a facilitator was spotting nearby cyclists using the mirrors and windows in the truck. Whenever a cyclist was spotted near the right side of the truck in a distance from 10 meters in front of the truck to 30 meters behind it, the facilitator would initiate the system using the interface on the tablet.

During the trials, the facilitator would ask questions and observe the truck driver's reactions. After each trial the facilitator would interview the truck drivers to go into depth with specific situations, and the truck drivers' overall impression of the system.

In general, the truck drivers were happy with the system, and the extra sense of security it gave. A few times the facilitator spotted the cyclist before the truck drivers, and started the playback. In these situations, the truck drivers looked towards the relevant mirror seeking visual confirmation and acted accordingly.

When talking about the sound the truck drivers would often use a phrase like: "When I heard the bike" or "I heard the bike coming from over there". This suggests that they had a mental model of the sound coming from the bike instead of the system.

The study indicated that not only could the truck drivers use the sound as an indication of nearby cyclists, they could also use it to guide their vision towards the correct mirrors.

The findings indicated that the truck drivers were pleased with the system and it seemed like they created a mental model of the sound coming from the cyclist. The system increased the awareness of the cyclists and helped the truck drivers find the cyclists in the mirrors.

9 Summary of Paper G: A Study of an Auditory Information System

Paper G first describes a listening experiment that examines which auditory icons that should be used to inform the truck driver that a cyclist is present near the truck. The research is used to decide on a type of sound stimuli to be used in the field study described in Paper F. Since the results of the field study is already described in Paper F it is left out of this summary.

Ten auditory icons were selected for a listening experiment. The auditory icons had different resemblance with a cyclist or a bike and were recorded binaurally using an artificial head designed at Aalborg University (Christensen, Jensen, and Møller 2000) so the spatial information was preserved. The auditory icons consisted of five different auditory icons each recorded in a static version and a version where the sound source was moved.

To create more realistic stimuli, the recordings of the truck cabin was mixed with the auditory icons. The sounds of a truck cabin were recorded binaurally by placing microphones in the ears of a truck driver at the blocked entrance of the ear canals. Two sessions were recorded: A) A session, where the truck driver was driving towards an intersection and made a right hand turn, and B) a session, where the truck driver drove straight ahead.

The recordings from the truck, and the recordings of the intended auditory icons was equalized and afterwards mixed together to form the experimental stimuli.

Twenty subjects participated in the experiment. None of the subjects had experience with driving a truck, but all had a driver's license for cars. The experiment consisted of two sessions. In both sessions the subjects should imagine that they were driving a truck. The first session (Session A) represented a critical situation where the truck driver is driving towards an intersection where the truck has to take a right hand turn. The other session (Session B) represents a trivial situation where the truck driver is driving straight ahead.

The experiment was carried out as a 2-alternative forced-choice. For each session the subjects were presented with all combinations of the 10 stimuli in pairs of two. For each comparison the subjects were asked to select the stimulus they preferred to represent a cyclist.

It was decided to use an indirect scaling methods to allow investigation into underlying attributes that affected the preference. The task of deciding which of two stimuli is greater, than the other is easier for untrained subjects, than expressing their perception or impression by means of an exact numerical value. The direct scaling methods means that it is uncertain, what type of scale the subjects actually use when rating the stimuli. The indirect scaling methods allow tests of consistency and scale type. The disadvantages of indirect scaling is the time it takes to complete all the comparisons and the number of subjects needed. Which in this case is low compared to recommendations (60) (Zimmer and Ellermeier 2003). The ratings you get from the indirect scaling methods are relative to the stimuli included in the experiment and not absolute.

The responses from the listening test were fitted to a BTL model (Tversky 1972) which can be seen in Figure 12. Due to a high number of strong stochastic transitivity violations in the second session, a preference tree was also used to fit the responses. The preference tree can be seen in Figure 13. The tree has a branching that represents the bell feature which was the most decisive attribute for the choice of stimuli. The model testing and fitting were done using a matlab function described in (Wickelmaier and Schmid 2004).

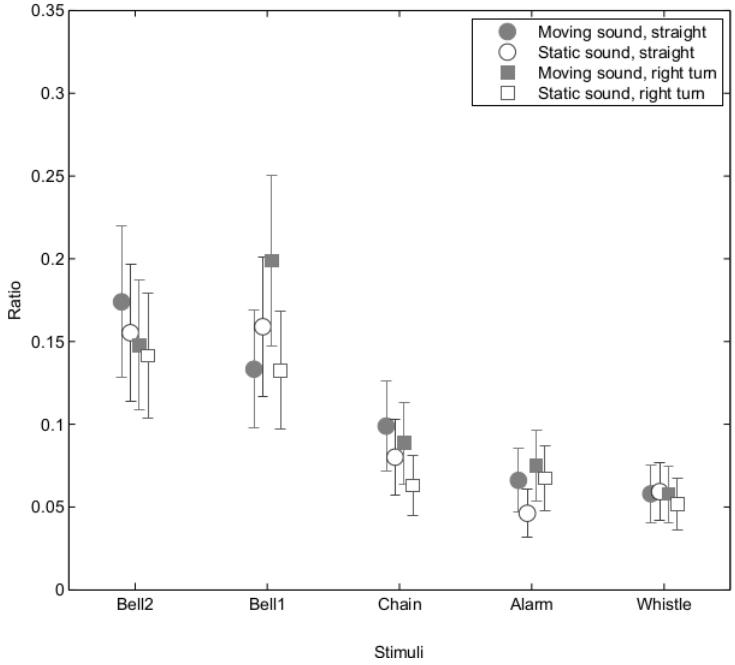

Fig. 12: The BTL model's values and confidence intervals for each stimulus for both sessions.

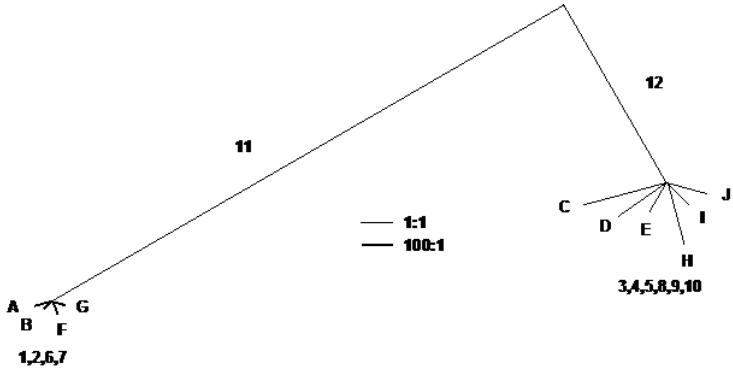

Fig. 13: A graphical presentation of the preference tree of the responses from the session B. The branches 11 and 12 represent the bell and "not bell" attribute. The ratio-scale value u for each of the stimuli is presented in Table 4 and the attribute value and standard error is presented in Table 5.

	Label	Stimuli	u-scale
Moving	B	Moving bell 1	0.3931
	A	Bell 2	0.3932
	C	Bike	0.1937
	D	Alarm	0.1748
	E	Whistle	0.1592
Static	G	Bell 1	0.3932
	F	Bell 2	0.3932
	H	Bike	0.1791
	I	Alarm	0.1595
	J	Whistle	0.1649

Table 4: u-scale value for each stimulus

Attribute	value	SE
1	0.000118	0.000421
2	6.05E-05	0.000216
3	0.05721	0.08256
4	0.03828	0.0513
5	0.02267	0.03359
6	9.60E-05	0.000341
7	0.000101	0.000361
8	0.04263	0.05982
9	0.0231	0.04059
10	0.0284	0.04059
11	0.393	0.1728
12	0.1365	0.1245

Table 5: The attribute value and standard error for the p-tree model

The results from the experiment show that bike bells, with no clear difference between the two types, are the most frequently chosen sounds for representing a bike moving along the right side of the truck followed by the sound of a bike chain, the alarm sound and the person whistling. The moving sound representations are generally chosen more frequently than the static versions.

In the laboratory experiment the background noise of the truck cabin was added to make it easier to imagine the situation but the experiment does not represent an ecologically valid situation, with the actual users of the system. It could be argued that a cyclist sounds like a cyclist no-matter the situation, but there is no way to know for certain. The use of truck drivers would have increased the ecological validity of the experiment. The results from the ex-

periment were, however, used to decide on what kind of auditory icon, that should be used for the 3D sound system, later tested in the field study. In the field study the truck drivers had a chance to experience the system in a scenario that came really close to the intended usage of the system. In the field study the truck drivers were pleased with the bell sound removing the concerns about the selection, based on the laboratory experiment.

Discussion

The idea with the design in Paper D was to create a challenging communication task where the business members' talk would overlap each other during the discussion but the use of only two members might not have been sufficiently challenging for the subjects to benefit from the binaural representation. Judged from the comments in the interview it seemed like the subjects were not conscious about the spatial properties of the sound introduced in some of the sessions.

The spatial properties might be more useful in a more challenging task with a larger auditory field and more sound sources or by having more overlapping speech. The speech-on-speech spatial release from masking test in Paper C provided a task where two similar sentences were spoken at once and showed an effect of using the hear-through compared to the occluded condition. This suggests that two sound sources are enough to make use of our spatial hearing but only if they are speaking at the same time. Another difference between the two experiments were that the subjects could use the facial expressions and lip movements in Paper D.

In the truck scenario the sound sources would appear outside of the visual field, and providing a different challenge in the sense that the auditory field is larger and the scenario is more dynamic with a moving listener and moving sound sources.

While the different scenarios weren't benchmarked against each other it seems that our spatial hearing proves more useful as the complexity of the task increases, which is by no means surprising.

In the different experiments I measured both the physical stimuli and the perception of these in different situations. in Paper A and B it was found that the hear-through device does not perfectly recreate the original sound. How these deviations affect our perception has to some degree been covered through the experiments in Paper C, but only in laboratory settings. The lab-

oratory experiments are good for benchmarking the system and to measure the limits of the system, but it does not necessarily say anything about how useful the system is in the specific application. How important is it to be able to localize with high resolution outside the horizontal plane for our ability to communicate in a teleconference application, and how do the constraints actually affect the experience of presence in a teleconference application?

In the experiment with the business meeting in paper D, and the pilot test in paper B with the puzzle, I tried to get closer to the actual application of the system, with the compromise of giving up some control of other variables. This gave new insights into the user's experience of the system, such as the subjects' experience of listening to their own voices or the background noise during silence.

As part of this ecologically approach I exclusively used untrained listeners, since expert listeners would have a different awareness of the auditory stimuli. One interesting finding that came up during some of the experiments was that the subjects wasn't always consciousness about the spatial properties of the sound, even in situations where they actually made use of them to perform better in the task.

Conclusion

The objective of the thesis was to identify ways of implementing 3D sound in selected multimodal applications and evaluate the feasibility of using 3D sound in these applications. The applications are; 1) a teleconference application and 2) a 3D sound system in trucks.

In the teleconference application a hear-through device was used to provide the locals with sound both from the other locals, and the visitor, and it was studied to what extend the spatial information could be preserved. The results from Paper A and B showed that the spatial information is preserved below 4-5 kHz. In Paper B and C it was shown that the hear-through device does not affect the communication quality and the perception of speech identification but do affect sound localization performance and the naturalness of the sounds.

The visitor was provided 3D positioned sound through headphones to represent the sound of the locals. The experiment in Paper D indicated that in the specific situation, with only two other persons sitting in front of the listener, the spatial information does not improve the communication.

In the 3D sound in truck application the test system presented in Paper E was used to provide the truck drivers with a 3D presentation of the cyclist. The truck drivers' responses in pilot test presented in Paper F and G indicated that 3D sound could be used to improve truck drivers' situational awareness regarding nearby cyclists.

The conclusion on the objective is that the implementation of the 3D sound in the truck improved the situational awareness of the truck driver, and can therefore be deemed feasible for the application. Implementations of 3D sound in teleconference application can have an effect as long as the task is sufficient challenging. The hear-through representation of the sound is sufficient for a communication situation like the teleconference application but may cause some limitations in other applications.

Future Research

In the hear-through design of the shelves solutions exist, where the microphone is build into the earphones, which could in theory bring it closer to the block entrance position, but the currently available solutions does not seem to perform better (Hoffmann, Christensen, and Hammershøi 2013a). This could however, be an option for future design of the hear-through device. Improvements could also be made by using individual equalizations, and by trying to lower the occlusion effect either by adding a leak or with active attenuation.

The performance of the hear-through device could further tested in a full sphere for a full assessment making it possible to conclude on the number of cone of confusion errors etc.

For the design of the truck system a study in a simulator could be used to further test the sound where the perceived urgency of the sound could be mapped to the situation. This could be adjusted by tuning attributes like amplitude, speed and fundamental frequency of the bells (Edworthy, Loxley, and Dennis 1991) using a rating method suggested in (Edworthy and Stanton 1995). In the same paper it is suggested to conduct an audibility test and a distinguishability test before the final design is evaluated. Because the sounds of the truck were part of the listening experiment it was possible to ask the subjects about any problems with audibility and distinguishability in the interview after the experiment. There could however be other sounds that was not included in the sound scenario. Furthermore, a more quantitative study of things like reaction time, arousal, situational awareness (Endslev 1996) and mental work load (Hart and Staveland 1988), could be carried out in a simulator for further assessment of the system.

In the final implementation of the system, a strategy for handling situations with several bikes. An option for user preferences like the volume of the system when the radio is turned on should also be incorporated in the design. The full implementation should be tested in a prolonged study of acceptance

(without a facilitator in the truck), and the effect should be measured (if it actually lowers the right-hand turn accidents).

References

Andersen, Richard A. et al. (1993). "Coordinate Transformations in the Representation of Spatial Information". In: *Current Opinion in Neurobiology* 3, pp. 171–176.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Third Edition. Wiley.

Begault, Durand R (1993). "Head-up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation". In: *Human factors* 35 (4), pp. 707–717.

Blauert, Jens (1997). *Spatial Hearing : The Psychophysics of Human Sound Localization*. Cambridge, Mass. MIT Press.

Bolia, R. S. (2004). "Special Issue: Spatial Audio Displays for Military". In: *The International Journal of Aviation Psychology* 14 (3), pp. 233–238.

Bolia, R. S. et al. (2000). "A Speech Corpus for Multitalker Communications research". In: *The Journal of the Acoustical Society of America* 107, pp. 1065–1066.

Bronkhorst, Adelbert W. (1995). "Localization of real and virtual sound sources". In: *The Journal of the Acoustical Society of America* 98 (5), pp. 2542–2553.

Bronkhorst, AW (2000). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions". In: *Acta Acustica United With Acustica* 86 (1), pp. 117–128.

Burkhard, M. and R. Sachs (1975). "Anthropometric manikin for acoustic research". In: *J. Acoust. Soc. Am* 58, pp. 214–222.

Christensen, Flemming, Clemen Boje Jensen, and Henrik Møller (2000). "The Design of VALDEMAR-An Artificial Head for Binaural Recording Purposes". In: *Audio Engineering Society Convention 109*.

Cruz-Neira, Carolina et al. (1992). "The CAVE: Audio Visual Experience Automatic Virtual Environment". In: *Commun. ACM* 35 (6), pp. 65–72.

Edworthy, J., S. Loxley, and I. Dennis (1991). "Improving auditory warning design: relationship between warning sound parameters and perceived urgency". In: *Human factor* 33 (2), pp. 205–31.

Edworthy, J. and N. Stanton (1995). "A user centred approach to the design and evaluation of auditory warning signals". In: *Human factor* 38 (11), pp. 2262–2280.

Endslev, M. R. (1996). "Towards a theory of situation awareness in dynamic systems". In: *Human Factors* 37 (1), pp. 32–64.

Fagerlönn, Johan (2010). "Distracting Effects of Auditory Warnings on Experienced Drivers". In: Washington.

Graham, R. (1999). "Use of Aditory Icons as Emergency Warnings: Evaluation Within a Vehicle Collision Avoidance Application". In: *Ergonomics* 42 (9), pp. 1233–1248.

References

Hammershøi, D. and H. Møller (1995). "Sound transmission to and within the human ear canal". In: *The Journal of the Acoustical Society of America* 100 (1), pp. 408–427.

Hammershøi, Dorte and Henrik Møller (2008). "Determination of Noise Immission from Sound Sources Close to the Ears". In: *Acta Acustica United With Acustica* 94 (1), pp. 114–129.

Hart, S. G. and L. E. Staveland (1988). "Development of NASA-TLX (Task Load Index) Results of empirical and theoretical research". In: *Advances in psychology* 52, pp. 139–183.

Hendrix, C. and W. Barfield (1996). "The Sense of Presence Within Auditory Virtual Environments". In: *Presence: Teleoperators and Virtual Environments* 5 (3), pp. 290–301.

Hoffmann, Pablo F., Flemming Christensen, and Dorte Hammershøi (2013a). "Insert earphone calibration for hear-through options". In: *Audio Engineering Society Conference 51th International Conference*. Helsinki.

— (2013b). "Quantitative Assessment of Spatial Sound Distortion by the Semi-Ideal Recording Point of a Hear-Through Device". In: *The Journal of the Acoustical Society of America* 133 (5), pp. 3283–3283.

Härmä, A. et al. (2004). "Augmented Reality Audio for Mobile and Wearable Appliances". In: *Journal of the Audio Engineering Society* 52 (6), pp. 618–639.

Huynh, H. and L. S. Feldt (1976). "Estimation of the box correction for degrees of freedom from sample data in randomised block and split-plot designs." In: *Journal of educational statistics* 1 (1), pp. 69–82.

ITU-T (2001). "ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ)". In:

Larsson, Pontus, Daniel Vastfjall, and Mendel Kleiner (2002). "Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering". In: *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society.

Larsson, Pontus et al. (2007). "When What You Hear is What You See: Presence and Auditory-Visual Integration in Virtual Environments". In: *Proceedings of the 10th Annual International Workshop on Presence*. Barcelona, pp. 11–18.

Levene, Howard (1960). "Robust Tests for Equality of Variances". In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Ed. by Ingram Olkin et al. Stanford University Press, pp. 278–292.

Macdonald, J. and H. McGurk (1978). "Visual Influences on Speech Perception processes". In: *Perception and Psychophysics* 24 (3), pp. 253–257.

Møller, H. et al. (1999). "Evaluation of artificial heads in listening tests". In: *J. Audio Eng. Soc* 47 (3), pp. 83–100.

Møller, Henrik et al. (1996). "Binaural Technique: Do We Need Individual Recordings?" In: *J. Audio Eng. Soc* 44 (6), pp. 451–469.

Møller, Henrik (1992). "Fundamentals of Binaural Technology". In: *Applied Acoustics* 36, pp. 171–218.

Møller, Henrik et al. (1995). "Head-Related Transfer Function of Human Subjects". In: *Journal of the Audio Engineering Society* 43 (5), pp. 300–321.

Perrott, David R. et al. (1990). "Auditory psychomotor coordination". In: *Perception and Psychophysics* 48 (3), pp. 214–226.

Rovstorn, P. (1982). "The W test for Normality". In: *Applied Statistics* 31, pp. 176–180.
Tikander et al. (2008). "An Augmented Reality Headset". In: *11th Conference on Digital Audio Effects*. Espoo.
Tikander, Miikka (2005). "Sound Quality of an Augmented Reality Audio Headset". In: Madrid.
Trafikstyrelsen (2014). *Strategi for Forebyggelse af Højresvingsulykker - mellem lastbil og cyklist*. Marts 2014. Rigspoliti, Trafikstyrelsen og Vejdirektoratet.
Tversky, Amos (1972). "Eliminition by Aspect: A Theory of Choice". In: *Psychological Review* 79 (4), pp. 281–299.
Veltman, J. A., Oving A., and A. W. Bronkhorst (2004). "3-D Audio in the Fighter Cockpit Improves Task". In: *The International Journal of Aviation Psychology* 14 (3), pp. 239–256.
Wickelmaier, Florian and Christian Schmid (2004). "A Matlab function to estimate choice model parameters from paired-comparison data". In: *Behavior Research Methods, Instruments, & Computers* 36.1, pp. 29–40.
Wilcoxon, F. (1945). "Individual comparisons by ranking methods". In: *Biometrics* 1, pp. 80–83.
Wright, S. P. (1992). "Adjusted P-values for simultaneous inference". In: *Biometrics* 48, 1005–1013.
Zimmer, K. and W. Ellermeier (2003). "Deriving ratio-scale measures of sound quality from preference judgments". In: *Noise Control Engineering Journal* 51 (4), pp. 210–215.

SUMMARY

This Ph.D. study has dealt with different binaural methods for presenting 3D sound in various applications. The purpose of this was to examine ways in which 3D positioned sound could be used in multimodal contexts. The impact and the feasibility of using 3D sound in the specific applications were studied.

The thesis dealt with a teleconference meeting where people interact in a virtual world. It was studied how changes in image quality and sound from two different sound rendering methods affect the communication in a teleconference meeting. In the teleconference meeting a system called a hear-through was used. The system allows you to capture the sound of your environment and to play it back through the earphones unattenuated.

The thesis also included work on a project with the goal of reducing the number of right-hand turn accidents by providing truck driver a 3D sound representation of cyclists as they approach the truck. A listening test is conducted with the purpose of studying which sound stimuli are best suited to represent a cyclist in the given situation.

The work was carried out at the Department of Electronic Systems, Aalborg University, Denmark in the period of 2012-2015.

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-399-9

AALBORG **UNIVERSITETSFORLAG**