Department of Computer Science and Engineering
Helsinki University of Technology
Espoo, Finland

Application of spatial sound reproduction in virtual environments – experiments in localization, navigation, and orientation

Matti Gröhn

Dissertation for the degree of Doctor of Science in Technology
to be presented with due permission of
the Department of Computer Science and Engineering
for public examination and debate
in Auditorium T1 at Helsinki University of Technology (Espoo, Finland)
on the 24th of May, 2006, at 12 noon.

CSC Research Reports R01/06
CSC – Scientific Computing Ltd., Espoo, Finland, 2006

ISSN 0787-7498
ISBN 952-5520-15-3 (printed)
ISBN 952-5520-16-1 (pdf)
Picaset Oy, 2006

HELSINKI UNIVERSITY OF TECHNOLOGY
P. O. BOX 1000, FI-02015 TKK
http://www.tkk.fi

ABSTRACT OF DOCTORAL DISSERTATION

Author Matti Gröhn

Name of the dissertation

Application of spatial sound reproduction in virtual environments – experiments in localization, navigation, and orientation

Date of manuscript 1.4.2006 | Date of the dissertation 24.5.2006

☐ Monograph ☒ Article dissertation (summary + original articles)

Department Department of Computer Science and Engineering
Laboratory Telecommunications Software and Multimedia Laboratory
Field of research Virtual reality, spatial sound reproduction
Opponent(s) Assistant professor Bruce Walker
Supervisor Professor Tapio Takala
(Instructor) Professor Tapio Takala

Abstract

The topic of this research was spatial sound reproduction in a cave-like virtual room (EVE) of the Helsinki University of Technology. Spatial sound reproduction is widely used, for example, in movie industry and computer games. In virtual environments it has been employed less than visual and tactile modalities. There are several common tasks in virtual reality applications in which spatial audio could be used. This thesis concentrates on localization, navigation, and orientation.

This research is one of the first studies in localization accuracy of loudspeaker reproduction in a virtual room. In the localization experiments, subjects pointed to the perceived direction of the sound source. According to the measurements, the achieved localization accuracy was at the same level as presented in literature for headphone reproduction. Localization of the moving sound sources was not as accurate as localization of the static sources.

In the navigation experiments, the task of the users was to move from waypoint to waypoint according to the visual and auditory cues. In the first experiment, auditory, visual, and audio-visual conditions were tested, and in the second experiment, different auditory cues were compared. Audio-visual navigation was the most efficient. Analysis of the travel paths indicated that an auditory cue was used at the beginning to locate direction of the next target, and a visual cue was used in the final approach to the target. In addition, all the subjects could navigate using the auditory cue alone. Auditory navigation performance increased when additional information about the distance and elevation of the target was included in auditory cues.

In the orientation experiment, subjects flew a predefined route inside an architectural model. Their task was to keep the model as balanced as possible during their flight. Three auditory artificial horizons were designed using “ball on a plate” metaphor. The sound was played from the direction towards which the virtual world was tilted. According to test results, the designed horizons helped the user to keep the model better in an upright position than without them.

Additional results included how the design of the virtual room and direction indication method affect on measured localization accuracy.

Keywords spatial sound reproduction, virtual reality, localization of sound sources, navigation, orientation

ISBN (printed) 952-5520-15-3 | ISSN (printed) 0787-7498
ISBN (pdf) 952-5520-16-1 | ISSN (pdf)
ISBN (others) | Number of pages 66 p. + app. 98 p.

Publisher CSC Scientific Computing Ltd.

Print distribution

☒ The dissertation can be read at http://lib.tkk.fi/Diss/2006/isbn9525520161/

TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi	VÄITÖSKIRJAN TIIVISTELMÄ
Tekijä Matti Gröhn	
Väitöskirjan nimi Tilaäänentoisto virtuaaliympäristössä - havaintoja paikantamisesta, suunnistamisesta ja orientaatiosta	
Käsikirjoituksen jättämispäivämäärä 1.4.2006	Väitöstilaisuuden ajankohta 24.5.2006
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)
Osasto Tietotekniikan osasto	
Laboratorio Tietoliikennohjelmistojen ja multimedian laboratorio	
Tutkimusala Keinotodellisuus	
Vastaväittäjä(t) Apulaisprofessori Bruce Walker	
Työn valvoja Professori Tapio Takala	
(Työn ohjaaja) Professori Tapio Takala	

Tiivistelmä

Tutkimuksen kohteena on ollut kaiuttimin tuotetun tilaäänentoisto Teknillisen korkeakoulun cave-tyyppisessä virtuaalitilassa (EVE). Tilaääntä käytetään laajalti muun muassa elokuvissa ja tietokonepeleissä. Keinotodellisuustutkimuksessa pääpaino on äänen sijasta toistaiseksi ollut näkö- ja voimapalautteen tuottamisessa. Keinotodellisuussovelluksissa on monia tehtäviä, joissa tilaääntä voidaan käyttää apuna. Tässä väitöskirjassan keskitytään paikannukseen, navigointiin ja asennon havaitsemiseen (orientaatioon).

Tämä tutkimus on ensimmäisiä joissa on mitattu virtuaalitilojen kaiuttimin toteutetun tilaäänentoiston paikantamistarkkuutta. Paikannustarkkuus mitattiin käyttäjätestein, joissa testihenkilöt osoittivat havaitsemansa äänen tulosuunnan. Mittausten mukaan saavutettu äänilähteiden paikannustarkkuus on samaa tasoa kuin parhaimmillaan kuulokkeita käytettäessä. Liikkuvien äänilähteiden paikantaminen on epätarkempaa kuin staattisten.

Navigointikokeissa tehtävänä oli liikkua kääntöpisteeltä toiselle annettujen vihjeiden perusteella. Kokeissa verrattiin ensin auditorisen, visuaalisen ja audiovisuaalisen vihjeen eroa ja toisessa vaiheessa erilaisten auditoristen vihjeiden eroa. Audiovisuaalinen navigointi oli kaikkein tehokkainta. Testihenkilöiden liikkumisreittejä analysoitaessa havaittiin, että alussa kohteen suuntaa haettaessa auditorisesta vihjeestä on kaikkein eniten hyötyä. Loppulähestyminen kohteeseen taas tapahtui parhaiten visuaalisen vihjeen avulla. Kaikki koehenkilöt suorituivat navigointitehtävästä, myös pelkän auditorisen vihjeen avulla. Kun auditorisen vihjeen avulla välitettiin myös etäisyys ja korkeustieto, käyttäjät navigoivat paremmin.

Orientaatiotestissä käyttäjien tehtävänä oli lentää ennalta määritelty reitti arkkitehtonisen mallin sisällä ja samalla pitää malli mahdollisimman vaakasuorassa. Testiä varten suunniteltiin ja toteutettiin kolme erilaista auditoristä keinohorisonttia. Testihenkilö sai kuulla auditorisen vihjeen suunnasta, johon malli oli kallistunut eniten. Testin perusteella malli pystyttiin pitämään paremmin vaakasuorassa auditoristen keinohorisonttien avulla kuin ilman niitä.

Päätulosten lisäksi tutkimuksen aikana saatiin selville, miten virtuaalitilan ja osoituslaitteen suunnittelu vaikuttaa mitattuun paikannustarkkuuteen.

Asiasanat Keinotodellisuus, tilaääni, äänilähteiden paikantaminen, suunnistaminen	
ISBN (painettu) 952-5520-15-3	ISSN (painettu) 0787-7498
ISBN (pdf) 952-5520-16-1	ISSN (pdf)
ISBN (muut)	Sivumäärä 66 s. + liit. 98 s.
Julkaisija CSC Tieteellinen laskenta Oy	
Painetun väitöskirjan jakelu	
☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/2006/isbn9525520161/	

to Päivi

Preface

This research was carried out in the Telecommunications software and multimedia laboratory (TML) at Helsinki University of Technology, Espoo during the years 1999 - 2005. My warmest thanks go to Professor Tapio 'Tassu' Takala, my Thesis supervisor, for his guidance and support during these years. I am indebted to all my co-authors Dr. Tapio Lokki, Professor Lauri Savioja, and Professor Tapio Takala. I am grateful to all my volunteer test subjects, the experiments could not have been accomplished without them.

I am grateful to Mr Juhani Forsman, and Mr. Seppo Äyräväinen for their efforts in keeping the EVE up and running. In addition, I thank Mr. Tommi Ilmonen for his efforts with EVE's spatial audio system. I am indebted to Mr. Iikka Olli for providing the software platform for the experiments. Co-operation with Laboratory of acoustics and audio signal processing has been crucial for this research. Especially I am grateful to Professor Matti Karjalainen, and Dr. Ville Pulkki.

Furthermore, I would like to thank the pre-examiners of my Thesis, Dr. Jim Ballas, and Dr. Nick Zacharov for valuable feedback, and constructive comments. In addition, I thank Ms. Carrie Turunen for her help in improving the English of this thesis. This study was financially supported by Helsinki Graduate School in Computer Science and Engineering (HeCSE), HPY Research foundation, and KAUTE foundation.

The research environment at TML was always enthusiastic and inspiring, for that I thank all the people there, especially Ursula Holmström, Timo Kiravuo, and Sanna Liimatainen. Several colleagues at CSC have supported me during these years. Especially I want to thank Juha Fagerholm, Pirjo-Leena Forsström, Leena Jukka, Minna Laine, Juha Ruokolainen, and Satu Tissari. Other special thanks go to current and former members of our in-house band Jussi Enkovaara, Anu Hämäläinen, Jaakko Korpela, Tommi Kutilainen, Roy Molini, Jouni Paltakari, Juha Ruokolainen, Peter Råback, Olli Serimaa, Tanja Tapio, Ismo Tossavainen, and Kalle Virta. Thursday evening sessions in Kaapeli have helped me to relax and forget the research and work.

My warmest thanks go to my parents Irja and Väinö Gröhn. They have always supported and encouraged me in my efforts. My brothers Timo, and Olli, relatives, and friends have reminded me, that the real life is outside the virtual environments. Thank you for you all.

Finally, it is time to express my loving thank to my wife Päivi, for her love, support, and encouragement throughout this long task. I am also grateful for my children Laura, Harri, and Ari. With my family I have found the meaning of my life.

Helsinki 1.4.2006

Matti Gröhn

PS. In addition to the Thesis, I have during this research also found out how many cans of vanSlooten is needed to accomplish one Doctoral Thesis. After careful empirical research the answer is 178.

Table of Contents

1 Introduction 9
1.1 Aim of the thesis . . . 11
1.2 Research questions . . . 11
1.3 Hypotheses . . . 12
1.4 Organization of the Thesis . . . 12
2 Spatial hearing and reproduction of spatial audio 13
2.1 Spatial hearing . . . 13
2.2 Spatial sound reproduction . . . 17
3 Related research in localization, navigation, and orientation 21
3.1 Localization . . . 21
3.2 Navigation . . . 24
3.3 Orientation . . . 26
4 Test environment and method 27
4.1 Virtual reality technology . . . 27
4.2 Research environment . . . 28
4.3 Method . . . 30
5 Results 34
5.1 Localization . . . 34
5.2 Navigation . . . 41
5.3 Orientation . . . 42
6 Discussion 46
6.1 Localization . . . 46
6.2 Navigation . . . 51
6.3 Orientation . . . 52
7 Summary and conclusions 53
7.1 Main results . . . 53
7.2 Applying the results . . . 54
7.3 Future directions . . . 55
Summary of publications and author's contribution 56
Bibliography 59
Errata 64
Original publications 66

List of Publications

This thesis summarizes the following articles and publications, referred to as [P1]-[P8].

[P1] M. Gröhn, T. Lokki, L. Savioja, and T. Takala. Some Aspects of Role of Audio in Immersive Visualization. In *Visual Data Exploration and Analysis VIII, Proceedings of SPIE Vol. 4302*, pages 13–22, San Jose, Jan. 2001.

[P2] M. Gröhn. Is Audio Useful in Immersive Visualization? In *Stereoscopic Displays and Virtual Reality Systems IX, Proceedings of SPIE Vol. 4660*, pages 411–421, San Jose, Jan. 2002.

[P3] M. Gröhn, T. Lokki, and T. Takala. Static and dynamic sound source localization in a virtual room. In *Proceedings of AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, pages 337–344, June 2002, Espoo, Finland.

[P4] M. Gröhn. Localization of a moving virtual sound source in a virtual room, the effect of a distracting auditory stimulus. In *Proceedings of the International Conference on Auditory Display (ICAD 2002)*, pages 394–402, Kyoto, Japan, 2.-5. Jul. 2002.

[P5] M. Gröhn, T. Lokki, and T. Takala. Localizing loudspeaker reproduced sounds in a cave-like room. *Presence: Teleoperators and Virtual Environments*. Accepted for publication in Presence Vol. 16, 2007, 19 pages

[P6] M. Gröhn, T. Lokki, and T. Takala. Comparison of auditory, visual and audio-visual navigation in a 3D space. In *Proceedings of the International Conference on Auditory Display (ICAD 2003)*, pages 200–203, Boston, 6.-9. Jul. 2003. Reprinted in ACM Transactions on Applied Perception Vol. 2 Nr. 4, pages 564–570, Oct. 2005.

[P7] T. Lokki, and M. Gröhn. Navigation with auditory cues in a virtual environment. *IEEE Multimedia*, Vol. 12, Nr. 2, pages 80–86, Apr.–Jun. 2005.

[P8] M. Gröhn, T. Lokki, and T. Takala. An orientation experiment using auditory artificial horizon. In *Proceedings of the International Conference on Auditory Display (ICAD 2004)*, Sydney, 6.-9. Jul. 2004., 6 pages

1 Introduction

This thesis is concerned with the application of spatial sound reproduction to virtual environments. Spatial hearing is an area that has been extensively explored, and is exhaustively covered in Blauert's book [1]. Historically, spatial sound reproduction has been developed from monophonic to conventional stereophonic to multichannel surround reproduction. In a monophonic system, the listener hears the sound from one point. In a stereophonic system, the sound source can be panned between left and right loudspeaker, and in multichannel surround systems, sound sources can be heard also behind the listener. In 3D sound reproduction, the listener also perceives the height of the sound source.

For several decades, spatial sound in movie theaters has been used for the sound effects and ambient sounds. Quite recently home theaters have become common in households, which has made spatial sound systems a part of everyday living. Another field in which the use of spatial sound is common is computer games. Most of the newest computer games have spatial sound included as a crucial part of the game experience. In games, spatial sound is used to provide information about target locations and background events. In addition, the player immersion with the game has been increased using carefully designed music and digital sound effects.

In the literature there are several different definitions for virtual reality and virtual environments. In this research, the definition by Kalawsky [2] is used: "Virtual environments are synthetic sensory experiences that communicate physical and abstract components to a human operator or participant. In a virtual environment, the human is immersed in a computer simulation that imparts visual, auditory and force sensations." Another sense-based definition of virtual reality is made by Burdea and Coiffet [3]: "Virtual reality is a high-end user computer interface that involves real-time simulation and interactions through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, smell and taste."

Sherman and Craig [4] use a subject oriented view in their definition: "Virtual reality is a medium composed of interactive simulations that sense the *participant's* position and actions and replace or augment the feedback with one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)."

So far the visual part has dominated in virtual reality research and spatial sound reproduction has been an underused feature. Although virtual reality is defined as multimodal, the audio is typically overruled by visual perception and even haptics. For example in Kalawsky's book [2] 95 pages describe visual displays, 15 pages for haptic devices, and 5 pages for spatial sound displays. In the book by Burdea and Coiffet [3] on virtual reality technology, the ratio is 27 pages for visual displays, 18 pages for haptic devices and only 8 pages for spatial sound displays. In the book by Sherman and Craig [4], the ratio is 48 pages for the visual display, 23 pages for haptic devices, and 13 pages for the spatial sound displays. In all three books most of the pages are used to explain the visual displays, and the least number of pages is devoted to auditory displays.

Pressing [5] has explored the different ways to use sound in virtual environments. He has divided sounds into three categories: artistic, informational, and environmental. Category 1 sounds include music and songs. It also includes virtual musical instruments and other auditory performances constructed in virtual environments. Category 2 sounds include speech, earcons [6], auditory alerts, and sonification [7]. It focuses on intended information transfer. Category 3 sounds include ambient sounds and sound effects. Ambient sounds are one of the most common ways to use auditory systems of virtual environments. These sounds are typically used to set the mood of the experience in a similar way to the film industry. Ambient sounds increase the mental immersion.

It is important that the auditory and visual cue are synchronized, if they are representing the same object. Hahn and Fouad [8] have explored integration of sounds and motions in virtual environments. When sounds and motions do not have the proper correspondance, the resultant confusion can lessen the effects of each other.

Spatial sound reproduction in virtual environments have been used to provide feedback on user actions. The difference between ambient sounds and interactive sounds is that ambient sounds do not directly respond to the user's actions. Interactive sounds in user interfaces have been explored for several years (see for example papers by Gaver[9, 10], Brewster [11], Mynatt [12], and Lucas [13]).

Virtual environments can be used for several different application areas. In flight simulators spatial sound reproduction can be applied to inform the pilot about the other planes and their locations. In addition, spatial sound reproduction increases the possibility to recognize different speakers in radio communication. In computer aided design process, the spatial sound reproduction could provide information about operation and status of the working machine. In architecture, spatial sound reproduction can be used to represent acoustics. In virtual environment games, the spatial sound reprodution can be used to localize targets or other important objects. In addition, ambi-

ent sounds and digital sound effects are used to get the player better engaged with playing experience. In scientific visualization the models are often complex and the spatial sound can be used to provide the user better insight to data. Sound can used to represent data or to localize the most important points/areas of the data.

In architecture, the most important thing is to reproduce as realistic sound field as possible. In games, scientific visualization, and flight simulators the directional accuracy is more important, and there is not as strong need for realism as in architecture.

1.1 Aim of the thesis

The aim of this research is to find out possible ways to apply spatial sound reproduction to virtual environments, or more precisely, in different virtual reality applications. There are several common tasks in these applications in which spatial sound reproduction could be employed such as localization, navigation, orientation, data representation, object selection, and object manipulation.

In many of these application areas the user is exploring large virtual worlds. To get insight of the world, the user should be able to localize different parts of the world. In large and complex virtual worlds, visual models might obscure each other, and spatialized auditory cues could provide additional help for the user to find the locations of interesting objects. In this research, localization is defined as the user's ability to define direction and distance of the target. Localizing objects is one of the most common tasks in virtual reality applications.

In large virtual worlds the user typically likes to move around. This makes navigation another common task in virtual worlds. In this research, navigation is defined as the user's ability to move from starting point to target. Spatial sound reproduction can provide some guidance to the user, which would make it easier to navigate. In orientation, auditory cues could be used to provide information about the different directions, and that way the user is better oriented and could, for example keep him/herself in an upright position.

1.2 Research questions

The first task in this research was to find out how well people could localize sounds in a virtual room. The next question was, can spatial auditory cues increase the navigation performance. Finally, could the spatial auditory cues provide better orientation awareness to users.

1.3 Hypotheses

These research questions lead to the following hypotheses, which were explored during this research.

- In a virtual room, it is possible to achieve a good localization accuracy with a loudspeaker reproduction comparable to a headphone reproduction.
- Additional distracting auditory stimulus decreases the localization accuracy.
- Auditory navigation is possible in a 3D environment even without any visual cues.
- Simultaneous visual and auditory cues support each other in navigation and it is more efficient than navigation using visual or auditory cues alone.
- Spatialized auditory cues improve the user awareness of the orientation.

1.4 Organization of the Thesis

Chapter 2 provides an overview of the spatial hearing and reproduction of the spatial sound. Related research in localization, navigation, and orientation are described in Chapter 3. Test environment and method are described in Chapter 4. In Chapter 5, the main results are represented. Chapter 6 discusses these results and compare them with results achieved in related experiments by other researchers. Chapter 7 summarizes and concludes this thesis.

2 Spatial hearing and reproduction of spatial audio

This chapter consists of two sections. The first section considers the main issues of spatial hearing and the second section covers the main methods used in spatial sound reproduction.

2.1 Spatial hearing

Typically, we associate a sound source with a direction and distance. In other words, we expect that most of the sound sources have a definite spatial location. Ambient sounds and sounds in reverberant environment have no definite location. An important difference between hearing and vision is that vision is limited on field of view, but hearing has no directional limitations. We can hear sound sources that are behind us or otherwise outside our field of view.

2.1.1 Ear canal signals

According to Blauert [1], the only information used in spatial hearing is the sound pressures in ear canals. These ear canal signals very seldom are the same as the sound signal emanated by a sound source. Over long distances air absorption affects the signal, ground reflects sound, and inside buildings, walls, ceiling, and the floor generate more reflections. One signal emanated from a sound source generates typically a group of sound signals that differ significantly from the original signal for example in their spectral content. In addition, they arrive at different times from different directions.

In addition to this, sound signals in ear canals change as a function of sound source direction relative to a listener. The arrival times of ear canal signals vary with direction because the ears are located on different sides of the head. In addition, the listener's head casts an acoustic shadow that causes spectral differences in ear canal signals. High frequencies are shadowed more than lower frequencies. In addition, pinna, head, and torso change the sound signal. These effects can be represented as free field transfer functions from a sound source to each ear canal. These functions are called head-related transfer functions (HRTFs) [1].

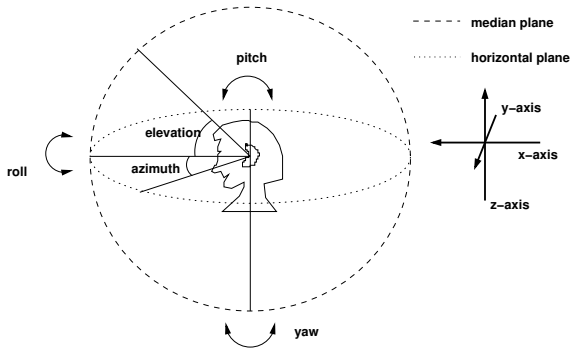

Figure 2.1: *Definitions of the median and horizontal planes. Azimuth angle is the angular difference from the median plane. Elevation angle is the angular difference from the horizontal plane. Yaw, pitch, and roll are the commonly used rotation angles around the z, y, and x coordinate axes respectively.*

In spatial hearing, listeners use different type of cues [14] such as: spectral content of ear canal signals and level, temporal or spectral differences between ear canal signals. Monaural cues are decoded from signals arriving in one ear, and binaural cues are derived from differences in ear canal signals.

2.1.2 Coordinate system for spatial hearing

In spatial hearing two important planes are the median and horizontal plane (see Figure 2.1). The median plane symmetrically divides the space relative to a listener into left and right parts. Thus for a sound source located in the median plane there is no difference between the sounds heard by each ear. Each point in the median plane is equidistant from both ears, and there is no arrival time difference for sound signals starting from median plane. The horizontal plane divides space into upper and lower parts. Each point in horizontal plane has the same height as ears.

The azimuth angle defines the angular difference from the median plane (in front of listener) and the elevation angle defines the difference from the horizontal plane. Directly in front of the listener, the value of the azimuth and elevation angle is zero. In this research, the azimuth angle is positive, if the source is located on the right of the median plane and negative if the source is located on the left of the median plane. Elevation values are positive when the source is above the horizontal plane and negative when the source is below the horizontal plane. In this thesis, x-axis is the front-back axis, y-axis is the left-right axis, and z-axis is the up-down axis. Roll is the rotation angle of the x-axis, pitch is the rotation angle of the y-axis, and yaw is the rotation angle of the z-axis.

In the following subsections, the main binaural and monaural cues in localization are presented. Other important concepts and factors relevant to

spatial hearing are also reported.

2.1.3 Binaural cues

There are two main directional binaural cues that are derived from differences in ear canal signals. Temporal difference is called interaural time difference (ITD) and level difference is called the interaural level differences (ILD). ITD is the primary cue in frequencies below 1.5 kHz and ILD is the primary cue above that threshold [1]. ITD and ILD are the strongest directional cues.

The ILD is small at low frequencies, regardless of source position because the dimensions of the head and pinna are small compared to the wavelengths of sound at frequencies below 1500 Hz [14].

2.1.4 Monaural spectral cues

Monaural spectral cues are important in elevation perception [14, 15]. The head, torso, and shape of the pinna affects spectral cues. Use of monaural spectral cues depends on both adequate high-frequency content in the auditory stimulus, and adequate high-frequency sensitivity on the part of the listener. Other monaural cues like temporal cues, and overall level are not important in the localization [14].

2.1.5 Precedence effect

The precedence effect [1] is a mechanism, that helps in the reverberant room to localize sound sources. In general, it is a psychoacoustic phenomenon whereby an acoustic signal arriving first at the ears suppresses the ability to hear any other signals, including echoes and reverberation that arrive about 40 ms after the initial signal, provided that the delayed signals are not significantly louder than the initial signal. The precedence effect searches actively for transients and uses binaural cues for a short time after transient has occurred. That way it ensures that localization is based on direct sound and reflections are not able to affect localization. The precedence effect can be interpreted as an "echo-avoidance" effect. According to Pulkki [15], listening test and simulation results achieved in anechoic conditions can be applied at least qualitatively in moderately reverberant conditions due to the precedence effect, because the direct sound is not changed by room acoustics.

2.1.6 Cone of confusion

The cone of confusion is an important concept in spatial hearing. It is defined as the set of all possible sound source locations, with the same time difference between ear canal signals [1]. For example sound sources directly in front, above or back the listener have the same time difference and without additional cues their locations could not be separate. Front-back confusions

are commonly reported in static localization experiments. The cone of confusion is effective in situations where the sound source keeps its position in the cone of confusion and the listener keeps their head still.

2.1.7 Head movements

Tilting of the head changes the monaural spectral cues and rotating the head changes the binaural cues. These changes can be used in localization. Head movements decrease the effect of the cone of confusion and improve the ability to determine the direction of the sound source [1].

2.1.8 Other factors

In addition to ITD, ILD and monaural spectral cues, there are also other factors involved in localization. Wightman and Kistler [14] have explored a variety of stimuli and listener factors, including stimulus dynamics, source familiarity, listener expectations, and cue plausibility. According to their article, the monaural spectral cues are less reliable than ITD and ILD. In addition, they mentioned that the use of monaural spectral cues depends critically on a listener's a priori knowledge of source characteristics. They also completed an experiment with unrealistic ITD or ILD cues. According to their results, the position judgments were always based on realistic cue, and listeners ignore the unrealistic cue.

According to Duda [16], the scale of the ILD spectrum determines azimuth, and the shape of the ILD spectrum determines elevation. This applies to the sources located closer than 2-3 meters from the head. This means that ITD and ILD alone could provide good accuracy in localization for near sources.

Middlebrooks [17] explained, that manipulations of the sound source spectrum, such as narrowband filtering, can generate erroneous localization responses, especially affecting elevation accuracy. When a stimulus contains a spectral peak, the localization judgments of subjects vary according to the center frequency of the peak. The perceived location tends to correspond the area of space for which the external ear most effectively collects sound at the frequency of the peak.

2.1.9 Moving sound source

According to Grantham [18], there seems to be two mechanisms in the human auditory system for the perception of moving sound sources. These mechanisms are a snapshot processor, that operates for rapidly moving events, and a motion sensitive mechanism that is tuned to slower velocity events.

In addition, Grantham [18] defined a minimum audible movement angle (MAMA). MAMA is the minimum angle that must be exceeded before it is possible to perceive the direction of motion of the moving sound source. Un-

der optimal circumstances (a slowly moving sound presented directly in front of the subject) the MAMA is between two to five degrees for the azimuth changes in the horizontal plane. This is larger than minimum audible angle (MAA) for the static sources, which is about one degree for low-frequency tones in front of the listener [18]. MAA is the minimum angle that must be exceeded before it is possible to perceive two static sound sources in different positions.

2.1.10 How to choose an adequate signal

An adequate signal for this research should be easy to localize and recognize. To use both main binaural cues (ITD and ILD), the signal should have enough energy below and above 1.5 kHz. In addition, it has been found [15] that frequencies near 6 kHz are important for elevation perception, at least while using vector base amplitude panning [19] in spatial sound reproduction. Different kind of noises are typically used in psychoacoustical experiments because as broadband signals, they cover the whole frequency area, and therefore activate both ITD and ILD.

In addition, the signal should include transients, that trigger the precedence effect. For example, noise bursts are better than continuous noise. In their experiments, Saberi and Hafter [20] found out that the interaural delay of the ongoing signal may affect the detectability of the direction of source movement. In addition, they reported that when two transients and spatially separated sounds occur within short temporal intervals ($<$100ms), a single sound image is perceived that continuously traverses through the spatial extent between the two sound sources.

In auditory signal design, one should be aware of the auditory illusions. For example, it is possible to generate the illusion of movement with static sound sources [21]. This situation occurs while using a brief amplitude modulation within a well-defined temporal window (and frequency channel) for a tonal stimulus under monaural conditions.

Finally, it would be more acceptable, if the chosen signal is not annoying to users. For example, in moving sound source localization experiments published in [P3-P5], almost all the subjects reported that the continuously repeating frog sound was irritating.

2.2 Spatial sound reproduction

In spatial sound reproduction, sound source could be produced in any location around the listener. Typically sound sources are divided in two types: real and virtual sound source. Real sound source is a physical sound source like a loudspeaker or a musical instrument. A virtual sound source denotes an auditory object that is perceived in location that does not correspond to any physical sound source [15]. Two main methods for spatial sound reproduc-

tion are the use of headphones or loudspeakers. These methods are explained more in detail in following sections.

2.2.1 Headphone reproduction

In headphone reproduction a sound signal is reproduced directly to the ears. It is possible to move the perceived location by manipulating the sound signal. Just changing the ITD values can move the perceived azimuth angle of the sound source location. Takala and Hahn [22] added a simple model for frequency independent ILD. This method is called the cardioid method, and it decreases front-back confusions.

More effective spatialization is achieved by using HRTFs. HRTFs are an intensively studied area in spatial sound reproduction. They can be divided into individualized, and non-individualized ones. Individualized HRTFs are measured individually for each listener. Often, individual measurements are not available and the non-individualized HRTFs should be used. Non-individualized HRTFs can be based on measurements of the other person ears or ears on the dummy's head. Non-individualized HRTFs are not as accurate as individualized HRTFs.

In headphone reproduction, a sound signal can be positioned in any direction if HRTFs for both ears are available [23]. Usually, digital filters are fitted to measured HRTFs [24]. In reproduction, these filters modify sound signals the same way as listener's torso, head, and pinna. If the listener'ss head moves during the listening, then this movement should be taken account in processing. This means that in a dynamic environment, a head-tracking and real time processing of HRTFs are needed for accurate reproduction. In virtual environments the user typically likes to move around, and the real-time update of the sound source positions is needed.

2.2.2 Loudspeaker reproduction

A variable number of loudspeakers can be used to implement 3D spatial audio. Theoretically 2 loudspeakers are enough for 3D sound reproduction. HRTF processing and cross talk cancellation [23] are both needed in this solution. This method has two main limitations 1) the best listening area (sweet spot) is very limited, and 2) it is critical to listening room conditions. The full spatial information can be retained only in anechoic chambers or listening rooms [24].

For home-theaters, the most common multi-loudspeaker system is 5.1 standard [25]. In a 5.1 system, there are five loudspeakers around the user and a subwoofer for the low frequencies. This is a good method to reproduce surround sounds provided in movie soundtracks, although it only provides horizontal direction.

In addition to home-theater systems, there are three commonly applied

multi-loudspeaker methods: wavefield synthesis (WFS), Ambisonics, and vector base amplitude panning (VBAP).

Wavefield synthesis

Wave field synthesis (WFS) [26] reconstructs a whole sound field. Theoretically, it is a perfect solution, but it is impractical in most situations. The WFS produces the sound field accurately only if the loudspeakers are at a maximum distance of a half wavelength from each other. For high frequencies, this distance is only a few centimeters, which cannot be achieved without a very large number of loudspeakers. In the horizontal plane using a 100 loudspeaker wave field can be produced spatially accurately for frequencies up to 1000 Hz. In this case, low-frequency ITD cues are produced accurately. Producing an accurate three dimensional wave field would need hundreds of loudspeakers and therefore WFS is not practical for 3D sound reproduction.

Ambisonics

Ambisonics is a recording and reproduction technique [27]. In reproduction, a special form of amplitude panning is used. A sound signal is applied to all loudspeakers with different gains, which could be negative or positive. To get an optimal result, the loudspeakers should be in a symmetric layout. Ambisonics can take place over a 360 degree horizontal only soundstage (pantophonic systems) or over the full sphere (periphonic systems). In 3D sound reproduction, Ambisonics is typically applied to eight loudspeakers in a cubical array, or to twelve loudspeakers as two hexagons on top of each other.

In Ambisonics, the directional accuracy degrades radically outside the sweet spot. The virtual sources are localized to the nearest loudspeaker that produces a signal. The signal arrive from the nearest loudspeaker first to the listener, and due to precedence effect, the virtual source is localized to the nearest loudspeaker. This takes place if the signal level of the nearest loudspeaker is not significantly lower (appr. 15 dB) than other loudspeakers producing the same signal [1].

Vector base amplitude panning

Vector base amplitude panning (VBAP) is a method used to calculate gain factors for pair-wise or triplet-wise amplitude panning [19]. The aim of the VBAP is to produce virtual sources, the direction of which is independendent of different loudspeaker setups. In 3D sound reproduction the virtual source is reproduced using the three closest loudspeakers at a time. When the number of the loudspeaker is greater than the three, an automated triangulation method [28] is used to define the loudspeaker triplets. If the listener moves away from the supposed listening position, then the direction accuracy does not decrease with VBAP as much as it does with systems that apply a signal

to all loudspeakers, such as systems that tune to the phase not only amplitude. VBAP enables arbitrary positioning of loudspeakers, which is a benefit in a virtual room because in a virtual room, visual displays limit the possible loudspeaker positions. In addition, VBAP is a computationally efficient method.

Summary

WFS aims to reproduce the complete wave field over a wide area but it is impractical in 3D sound reproduction. Ambisonics and VBAP attempt to reproduce the sound at a certain listening point, and they both are suitable for 3D sound reproduction. VBAP is less sensitive to listening position and it allows a more arbitrary loudspeaker positioning than Ambisonics. In our research environment the VBAP was already installed before this research therefore, VBAP was used in this study. The choice is justified, because Ambisonics would not provide more accurate spatial sound reproduction than VBAP, and WFS is impractical in 3D sound reproduction.

3 Related research in localization, navigation, and orientation

In this chapter, findings from other researchers in localization of the sound sources are provided. Next, experiments in navigation in virtual environments and navigation using auditory cues are reviewed. Finally, results on orientation in virtual environments are presented.

3.1 Localization

Localizing real sound source in an anechoic environment is exhaustively covered in [1]. Blauert defined that localization is the law or rule by which the location of and auditory event (e.g., its direction and distance) is related to a specific attribute or attributes of an auditory event. Localization blur is the amount of displacement of the position of the sound source that is recognized by 50 percent of the experimental subjects as a change in the position of the auditory event. Accuracy of the spatial hearing is the amount of the localization blur under optimum conditions. In front of the listener, localization blur in the azimuth is approximately one to four degrees depending on the signal. In elevation, the localization blur is more signal dependent and it can vary from four degrees (white noise) up to seventeen degrees (speech by an unfamiliar person).

According to Blauert [1], the distance of the sound source, and distance of the perceived auditory event corresponds quite well for familiar signals such as human speech at its normal loudness. For unfamiliar sounds, localization with respect to distance of the sound source is largely undefined. Typically, localization experiments cover only the direction measurements (including experiments reported in publications [P2-P5]).

Auditory localization of static virtual 3D sound sources has been previously tested in several experiments. Most of these tests have used headphone reproduction with HRTFs [29, 30, 31, 32, 33]. As a part of their localization experiments, many researchers and research groups like Sandvad [34] and Martin et al. [35] have measured the localization accuracy of a real source

(loudspeaker) in an anechoic chamber. The localization accuracy of a virtual source in amplitude panned loudspeaker reproduction has been reported, for example, by Pulkki [36, 37, 38]. According to the author's knowledge, localization experiments have not been accomplished in cave-like virtual rooms prior to the experiments described in articles [P2-P5].

Comparison of results of different localization experiments is not straightforward for several reasons. Experimental designs, and methods differ. In some experiments, head movements are allowed and in others they are not. Different signals are used in experiments, and according to real source experiments [1], they provide different level of accuracy especially in elevation.

Results are reported using azimuth, and elevation errors, or the combined error angle. Azimuth and elevation errors are the difference between sound source azimuth and elevation angles and measured azimuth and elevation angles of the pointing direction. The following formulas are used to calculate the error angle:

$$\begin{aligned} x_s &= \sin(azi_s) * \sin(ele_s) \\ y_s &= \cos(azi_s) * \sin(ele_s) \\ z_s &= \cos(ele_s) \\ x_m &= \sin(azi_m) * \sin(ele_m) \\ y_m &= \cos(azi_m) * \sin(ele_m) \\ z_m &= \cos(ele_m) \\ Error\ angle &= \arccos(x_s * x_m + y_s * y_m + z_s * z_m), \end{aligned}$$

where azi_s is the azimuth angle of the source, and ele_s is the elevation angle of the source measured as in Figure 2.1. Corresponding angles for the measured direction are azi_m and ele_m. $[x|y|z]_{[s|m]}$ are cartesian coordinates on a unit sphere for the source and measured direction. $Error\ angle$ is the shortest angular distance on a unit sphere between the sound source and the measured location that is the angle between the vectors pointing from the listener to the sound source and its perceived location respectively.

In localization experiments, a direction indication method provides information about the perceived location. Researchers have used several different direction indication methods like graphical response screen [39, 33], adjusting a reference sound [40, 37, 38], pointing on a schematic drawing of the loudspeaker setup [41], a head mounted laser pointer [35], nose pointing [17, 42] or a tracked toy gun [34].

Djelani et al. [32] have compared three different direction indication methods: the Bochum-Sphere technique (also known as GELP), finger pointing and head pointing. In the Bochum-Sphere technique the position of the

auditory event is indicated on a sphere representing auditory space. According to their results, finger pointing and head pointing were superior to the Bochum-Sphere technique for localization experiments.

In experiments reported by Djelani et al. [32], they compared the effect of the head movements in localization accuracy. They used individualized HRTFs in reproductions and white noise as a stimulus. According to their results, allowing head movements significantly increases the localization accuracy.

Begault et al. [33] evaluated the effects of the following variables in accuracy: head movements, individualized HRTFs, and early and diffuse reflections. According to their results, the inclusion of reverberation decreased azimuth error and increased elevation error. An interesting result is that with speech signal, individualized HRTFs and allowing the head movements did not improve the localization accuracy. Authors emphasized that results may differ when broadband stimuli are used.

In his experiment Sandvad [34] drew the following conclusion on localization of virtual and real sources: Subjects localized virtual sound sources almost as well as real sound sources, but the increase in elevation error was statistically significant. Virtual sources were reproduced using headphone reproduction with HRTFs.

Martin et al. [35] achieved the same localization accuracy with real sources and virtual sources. In their experiment, they had only three subjects and they used gaussian noise as the broadband auditory stimulus. As well, they used headphone reproduction with the HRTFs for the virtual sources.

In their stereophonic panning paper, Pulkki and Karjalainen reported [36], that subjects could localize most natural broadband signals to an intended direction with a fairly good accuracy. According to their results, localization of broadband signals is defined mostly by low-frequency ITD and high-frequency ILDs. If a virtual source contains frequencies only above the 1100 Hz, the localization will be based mostly on ILD cues, which makes the virtual source direction more spread. In addition, the perceived direction varied between the subjects.

In a follow up paper to [36] Pulkki explored the localization of panned virtual sources in two-dimensional and three-dimensional situations [37]. According to results stimulus should include high-frequency components near 4 kHz to ensure good accuracy in elevation perception. With the broadband stimulus (pink noise) the subjects consistently perceived the azimuth direction. Perceived elevation directions of amplitude-panned virtual sources varied between subjects.

In virtual environments there typically are delays in communication between the tracking system, application, and spatial sound reproduction. The effect of the latency in localization of HRTFs reproduced sound sources is explored for example by Wenzel [39]. According to results elevation con-

fusions and error angles increased with latency, although the increases were significant only for the largest latency tested, 500 ms. Wenzel reminded, that other tasks, such as tracking an auditory-visual virtual object, may be much more sensitive to latency effects.

3.2 Navigation

In the literature navigation is often divided into two subtasks, way-finding and travel [4]. Way-finding is knowing where you are and how to get where you want to go. Travel is the act of moving through a space.

Different aspects of navigation in virtual environments have been actively explored. For example, Steck and Mallot [43] explored the role of local and global visual landmarks in virtual environment navigation. According to their results, some of the subjects used only local landmarks while others relied exclusively on global landmarks. Some of the subjects used local landmarks in one location and global landmarks at the other. This experiment used visual landmarks. Auditory beacons could be used as auditory landmarks.

Ruddle and Payne [44] have investigated components of subjects' spatial knowledge when they navigated large-scale virtual buildings using a non-immersive virtual environment. According to their results, familiarity with the model was the most important factor in subjects navigation performance. However when a compass was given to subjects, there was no increase in their performance, which was a little bit surprising.

Chen and Stanney [45] proposed a theoretical model of wayfinding that can be used to guide the design of navigational aiding in virtual environments. They divided navigation in three subprocesses: cognitive mapping, wayfinding plan development, and movement through an environment. In addition, taxonomy of navigational tools was proposed that would divide them into five functional categories: 1. tools that can display an individual's current position, 2. tools that can display an individual's current orientation, 3. tools that can log an individual's movements, 4. tools that can demonstrate the surrounding environment, and 5. guided navigational systems. In the real world, GPS receivers can take care of categories 1, 2, and 3.

In the following experiments, auditory cues in navigation are explored. According to the navigation tool taxonomy these all are experiments with category 5 tools. Auditory cues can provide information about the locations, that are not in the field of view or otherwise not visible.

One of the first reported experiments using spatial auditory display in virtual reality navigation experiment is by Darken and Sibert [46]. In their navigation experiment, they used the AudioCube eight loudspeaker system to provide spatialized virtual sources. They were added to the start location as a cue for the homing task. The audio signal was used for rough direction

finding.

Loomis et al. [47] focused on the development of a navigational system for visually impaired people. They evaluated guidance performance in a 2D navigation experiment as a function of four different display modes: one involving spatialized sound from a virtual acoustic display, and three involving verbal commands issued by a synthetic speech display. In this experiment spatialized sound was better than any of the verbal command methods.

AudioGPS [48] was designed to be a minimal attention user interface for the sighted person who is simultaneously involved in other demanding tasks. Direction was indicated by amplitude panning a virtual source. To avoid front-back confusion a different tone was used for the front and back directions. In addition, the angular difference between the travel path and destination were sonified using pitch difference. Distance information were sonified using Geiger counter metaphor. Tempo of the pulses of sound increased as a destination was approached. AudioGPS was found suitable for locating target on foot, but it was too slow in response to rapid changes of direction to be suitable for car driving.

Optiverse (Sherman and Craig [4], page 340) uses distance sonification to guide the user. Two sounds of the same timbre are emitted, and they differ in frequency based on the distance from the destination.

Rutherford et al. [49] proposed the use of auditory beacons to aid emergency egress from buildings, ships, oil exploration platforms and aeroplanes.

Navigation is typically explored using task based experiments. Subjects travel to specific locations or follow a predefined route. In the following experiments, these specific locations and route waypoints are marked using an auditory cue.

In our preliminary 2D navigation experiment headphones were used [50]. In this experiment, we compared three different stimuli, three different panning methods, and three different acoustic conditions. According to our results, pink noise provide the better navigation performance than artificial flute and guitar. In addition, anechoic condition provides better performance than reverberant conditions.

Walker et al. completed two 2D navigation experiments [51, 52] using headphone reproduction with non-individualized HRTFs. In their experiment, the subject's task was to walk a predefined route. The route was marked using auditory waypoints. One waypoint at a time was audible. The capture radius is a radius around the waypoint that is considered close enough so that a current waypoint could be muted and the next waypoint can appear. They have found that the type of auditory cue affects navigation performance [51]. According to their results, noise burst provides better performance than pure tone and sonar pulses.

In their follow up experiment [52], Walker et al. explored the effect of the size of the capture radius using the same set of the auditory cues as in their

previous experiment. In their experiment, the medium size radius was the slowest but the normalized path lengths were the shortest. The largest radius was the fastest, but subjects travelled longer path lengths. With the smallest radius, subjects travelled the longest path lengths because the subjects had some difficulties finding the waypoints. The smallest radius was faster than the middle radius, and slower than the largest radius.

3.3 Orientation

Bowman et al. [53] explored the effects of various travel techniques in an immersive virtual environment on the spatial orientation of the users. Spatial orientation here means the sense of position and orientation. Although navigation is divided into two parts in the literature: way-finding, and travel, people do not separate their navigation in this way, and the method of travel may have an effect on spatial orientation. According to their results, subjects performed better in 2D conditions than in 3D conditions. Travel techniques generated no significant differences.

According to Lackner et al. [54], spatial orientation is a substantial component for the sense of presence in real and virtual environments. Thus, body position in relation to the structural features and dependencies of the spatial surroundings have a profound influence on perceived spatial orientation. In virtual environments, it is possible to create visual perspectives that would be impossible under terrestrial conditions. This kind of perspective could give rise to illusory changes in self-orientation and visual orientation. These illusions should be taken account when a virtual environment is used to train spatial knowledge of a real environment.

According to the author's knowledge, there is no published auditory orientation experiments in virtual environments available.

4 Test environment and method

In this chapter, virtual reality technology in general, and the research environment are first presented. Next, the scheme for task based auditory experiments is presented. The last part of this chapter is used to describe how this scheme is applied in localization, navigation, and orientation experiments.

4.1 Virtual reality technology

To achieve the immersion described in Chapter 1, different display technologies and interaction devices have been used. Methods for spatial audio reproduction were covered in Chapter 2.

Virtual reality technology is not the new invention as people commonly believe. Forty years ago, Ivan Sutherland built Ultimate Display [3], which consisted of two cathode ray tube (CRTs) mounted along the users ears. Due to the weight of the display, Sutherland supported the display using a mechanical arm, which was also used as a mechanical tracking system.

Burdea and Coiffet [3] divided visual displays into two classes according to number of users: personal displays, and large-volume displays. Latter displays allowed for several users. Head-mounted displays (HMDs), hand-supported displays, floor-supported displays, and autostereoscopic monitors belong to personal visual display category. Large-volume displays could be divided to monitor based (single or side-by-side CRTs) and projector based displays like workbenches, CAVEs [55], display walls, and domes. Spatial sound reproduction with headphones (see section 2.2.1) is especially suitable to use with HMDs, because headphones are easy to integrate with HMDs, and head-tracking is typically included. Loudspeaker reproduction (see section 2.2.2) is suitable with large-volume displays, especially with cave-like environments, because they are planned to be multiuser environments.

Most of the interaction in the virtual environment depend on knowledge of real-time position and orientation of the moving objects like interaction devices in the user's hand. A tracker is a device that provides this information to the virtual reality applications. There are several alternatives for tracking like mechanical, magnetic, ultrasonic, inertial, and optical trackers. Wireless trackers are the most suitable option but they are also the most expensive solution. Mechanical trackers are impractical for cave-like environments,

because they are hard to implement inside them.

Experiments reported in this thesis were accomplished in a cave-like environment and are explained more in detail in the following subsection.

4.2 Research environment

All experiments were done in a virtual room of the Helsinki University of Technology called EVE [56]. EVE is a cave-like virtual room with four visual displays, three walls and a floor, as seen in Figure 4.1. In EVE there are fourteen Genelec 1029A loudspeakers and one subwoofer for spatial audio reproduction. Virtual sound sources are reproduced using VBAP.

Most of the loudspeakers are located behind the visual display screens as seen in Figure 4.1. The screen between the loudspeaker and listener affects the perceived sound signal. Measurements in EVE have shown that high frequencies of direct sound are attenuated more than 10 dB. Compensation filters are used to equalize the frequency response. Detailed descriptions of the screen compensation and implementation details of the audio reproduction system are described in articles [57, 58].

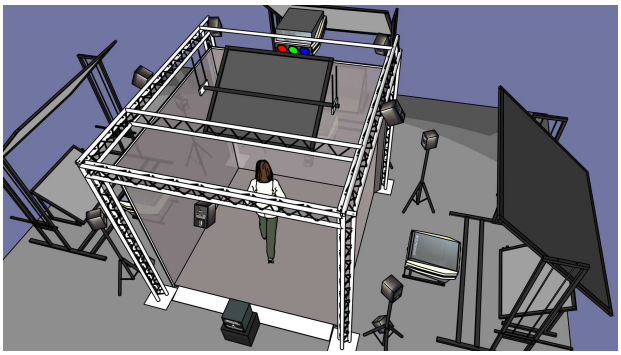

Figure 4.1: *Schematic drawing of the EVE (courtesy of Seppo Äyräväinen).*

For tracking the direction indication device, an Ascension Motionstar magnetic tracker was used. According to the manufacturer's technical specification [1], its accuracy (see Table 4.1) is as high or even higher than accuracy of human auditory localization [1], and therefore it is good enough for the experiments. In localization experiments a tracked baton was applied as a direction indication method and a custom made wand as an interaction device. The wand consists of a radio mouse (Locitech Surfman) and a tracker sensor. In navigation and orientation experiments, the wand was used to control direction and velocity of the movements. The gesture of pushing a wand but-

[1] http://www.ascension-tech.com/products/motionstar.php (visited 10th of October 2005)

Table 4.1: *Accuracy of the Ascension Motionstar magnetic tracking system according to the manufacturers technical specification.*

	1.52 m range	**3.05 range**
Position		
Static accuracy	0.76 cm	1.50 cm
Static resolution	0.08 cm	0.25 cm
Orientation		
Static accuracy	0.5^o	1.0^o
Static resolution	0.1^o	0.2^o

ton and moving the wand in space defines vector, the length and direction of which are translated into motion speed and direction in a virtual space. The turning of the wand control the rotations.

4.2.1 Changes in environment during the research

After experiments described in the articles [P2-4], a systematic error in elevation measurements was found. The measured elevation was consistently above sound source elevation.

Due to this systematic error in elevation perception, the localization for different directions was explored more in detail with informal testing using the following method. A set of different stimuli (high- and low-pass filtered noise and some filtered speech signals) that could be interactively changed was used. The virtual sound source was moved interactively using our tracked baton as an auditory pointer. Using this interactive method, comparison and investigation of different sound source direction stimulus combinations was efficient. With this informal listening test, it was found that elevation perception might have been distracted by quite loud ceiling reflections. Due to this finding, sound absorbing material was added above the screens to diminish the effect of those reflections.

In addition, it was found that the original direction indication method with the tracked baton produced elevation error of about five degrees due to the way the user held the baton in the experiment. To minimize error, the tracked baton was equipped with a handle and a sight.

The third possible source for systematic error was the calibration method used to adjust and equalize the levels of each loudspeaker. Previously, white noise was used as a calibration signal. With that signal, room modes at low frequencies might have affected the calibration. To minimize the effect of the room modes, the calibration signal was changed to narrow-band noise near 1000 Hz.

After the three step enhancements, the environment was measured again and no systematic error was found.

4.3 Method

Since the goal is to determine how well and accurately the user can perform in different tasks, task based user tests were chosen as a research method. In localication experiments, the subject's task was to point to the direction of the perceived location of the target sound source. In navigation experiments, the subject's task based action was to fly through a predefined track. In the orientation experiment, the task-based action was to fly a predefined route inside an architectural model. During the route, subjects needed to keep the model as upright position as possible. Subjects had a training period, which included one or more subtasks, before the experiments. In all experiments, the order of the subtasks was randomized to avoid the learning effect. In addition, in all experiments the subjects were asked to stay in the middle of the EVE. This was controlled by the test conductor.

According to Zahorik et al. [59], there is no significant difference in localization performance between the anechoic virtual world and echoic virtual world. However, in our preliminary navigation experiment [50], we found that the anechoic situation was most suitable for navigation tasks. As a result, it was decided that all experiments would be conducted in an anechoic virtual world.

Figure 4.2: *Scheme for task based auditory experiments.*

There are five factors affecting the results of the auditory experiments as seen in Figure 4.2. Each of these factors affects the measured results. If the signal does not provide enough cues, then its localization is inaccurate. Broadband signals such as pink noise use both main binaural localization cues (ITD and ILD). In a virtual room, the screen between the loudspeaker and listener produces a coloration of the perceived sound signal. In addition, the virtual room has its own acoustics and reverberation. The reproduction could have reflections that shift the perceived location of the stimulus.

Virtual sources are perceived more inaccurately than real sources. Localization accuracy of the auditory cues is, at best, one degree in azimuth angle and four degrees in elevation angle [1]. Subject's task based actions provide their own inaccuracy, and in addition, all the measurements are dependent on the accuracy of our tracking system. For the task based actions, the wand and the tracked baton described in section 4.2 were used. In virtual environments, there are many other possible interaction methods and devices, but

these were not in the scope of this research.

4.3.1 Localization experiment

There were three different localization experiments: static sound source [P5], moving sound source [P5], and moving sound source with distracting auditory stimulus [P4]. All localization experiments were accomplished without any visual cue. In the static sound source and moving sound source experiments described in [P5], there were eight non-paid volunteers. In the moving sound source experiment with distracting auditory stimulus described in [P4], there were also eight non-paid volunteers. Six of them were the same as in the other experiments described in [P5]. In all experiments, each of the subjects reported having normal hearing, although this was not verified with audiometric tests.

In the static sound source experiment [P5], a pink noise burst played in a continuous loop was used as a stimulus. Total length of the burst was 820 ms with 20 ms attack and release time. There was 30 ms break between the bursts. Guidelines for selecting an adequate stimulus are presentend in section 2.1.10,

In the moving sound source experiments [P4,P5], there were three different stimuli: pink noise (one minute long sample), music (2 minutes 45 seconds long excerpt from The Wall by Pink Floyd) and a frog croak (0.5 second long). In the experiments, the stimuli were played continuously in a loop. These signals had a different spectral content. Pink noise covered the whole audible frequency range without temporal structure. Music was a broadband signal that had a clear temporal structure. The croak sound had most of its energy below 2 kHz.

In all these experiments [P4,P5], the measured azimuth and elevation values for perceived location were recorded as well as the azimuth and elevation values for the sound source location. In the static sound source experiment, the target stayed in one position, and in the moving sound source experiments subjects followed the movement of the target by pointing to its perceived location with the pointing device.

During the experiment, the subjects could freely turn their head and body. The subjects pointed to the perceived location of the target sound with a pointing device and indicated it by clicking a button. In the static sound source experiment [P5], after the subjects clicked the button, the sound was muted. The next sound was played after a short pause.

In the moving sound source experiment [P5], after the user clicked the button, a signal was heard that indicated that the target sound had started to move. The subject's task was to follow the movement of the perceived sound source by pointing to it with a pointing device. The end of the movement was indicated using an end signal. There was a short pause before the next sound

was played.

In the moving sound source with the distracting auditory stimulus experiment [P4], the target sound was first played in a starting position without the distracting stimulus. After the subjects clicked the button of the wand, they heard a signal that indicated that the target sound had started to move. Simultaneously, the additional distracting sound started. In this experiment, the distracting stimulus was always the same as the target sound but with different timing and gain. The gain of the distracting sound was 10 dB less than the gain of the target sound.

In auditory localization experiment, the direction indication method should be at least as accurate as human auditory perception. In experiments described in [P2-P4], a tracked baton was used as a pointing device. In experiments described in [P5], a handle and sight was attached to a tracked baton for more accurate pointing.

4.3.2 Navigation experiment

A game-like experience was designed for the navigation experiments [P6, P7]. Subjects flew through a predefined track that was located inside a large, visible protein-drug complex. Track corners, called gates, were indicated with visual and/or auditory stimulus. Each run started in the middle of the virtual world and the first gate of the track was randomly chosen. Each run lasted a fixed time and the number of found gates was recorded.

Two navigation experiments were accomplished: a comparison of auditory, visual, and audio-visual cues [P6] and a comparison of different auditory cues [P7]. In the first experiment [P6], the visual cue was a white ball, the auditory cue was pink noise bursts with $1/r$-law distance attenuation (thus this stimulus is called *Gain* in the next experiment [P7]) and the audio-visual cue combined the visual and auditory cue.

In comparison to different auditory cues [P7], there were four different auditory stimuli and no visual cue. The first auditory stimulus *Gain*, was the same used in the audio-visual test. In the second signal *Gain+Rate*, the distance to the next gate was indicated with a density of bursts. The third signal *Gain+Pitch* consisted of noise bursts and a narrow-band noise, the center frequency of which represented the height of the sound source. The fourth signal, *All*, combined all cues.

In comparison to the auditory, visual and audio-visual cue [P6], there were nine non-paid subjects and for the comparison of different auditory cues [P7], there were eight non-paid subjects.

4.3.3 Orientation experiment

There were eight non-paid subjects in the orientation experiment [P8]. In the orientation experiment, subjects flew a predefined route inside an archi-

tectural model. During the route, subjects had to keep the model as upright as possible. An auditory artificial horizon was designed for this experiment [P8]. An obvious way to use 3D audio for orientation information is to mark the ends of the coordinate axes with auditory beacons in front and on the side. When both beacon sounds are heard at the ear level, both roll and pitch angles are close to zero. However, since elevation perception is not very accurate, this was found to be impractical in our informal tests. In addition, as Benson [60] discussed, a sound source, fixed with respect to the observer, does not give an intuitive feeling of orientation.

A better way to indicate disorientation was found by applying a "ball on a plate" metaphor; when the plate is tilted i.e. deviated from the upright position, the ball starts to roll to the direction pointing downwards. This metaphor was applied to the 3D auditory display so that sound was heard from the direction tilting downwards. In fact, with this metaphor the elevation information (spatial disorientation) was mapped to the azimuth angle. From the point of view of human spatial hearing, this mapping is more optimal since the azimuth perception is more accurate than the elevation perception [1].

Three different auditory stimuli were applied for the auditory artificial horizon in this experiment [P8]. All auditory stimuli were based on pink noise bursts. In the first stimulus, the amount of tilt was used as a gain factor. When the model was oriented, the stimulus was inaudible.

In the second stimulus, the pulse rate of the noise burst was varied according to amount of tilting. If the model was oriented upright, the rate was 0.7 Hz. The maximum rate was 8 Hz. In other cases, the rate of the stimulus was 2.4 Hz.

In the third case, a narrow band-pass noise was added to the stimulus. The center frequency of the noise varied from 50 Hz (when oriented) to 2 kHz.

5 Results

In this chapter, the main results of the localization, navigation, and orientation experiments reported in [P2-P8] are presented.

5.1 Localization

In analysis, azimuth, elevation and error angle as defined in section 3.1 were used. Localization results are presented using median error [1]. The median is a robust estimate of the center of a sample of data, since outliers have little effect on it. The main results are from experiments described in [P4] and [P5].

5.1.1 Direction indication method

As a part of the static source experiment described in [P5], the pointing accuracy of the direction indication method was measured using ten visual targets (covering the visual display area) for each subject. These measurements were taken before the auditory stimulus experiment.

[1] for example median elevation error = median(abs(measured elevation - source elevation))

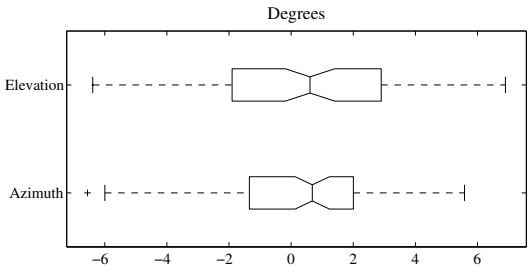

Figure 5.1: *Box plot of the error in visually guided response using the device used in the localization experiment. The left and right lines of the "box" are the 25th and 75th percentiles of the sample. The line in the middle of the box is the sample median. If the median is not centered in the box, that is an indication of skewness. The "whiskers" show the extent of the rest of the sample. The notches in the box are a graphic confidence interval about the median of a sample. A side-by-side comparison of two notched box plots is the graphical equivalent of a t-test.*

Figure 5.1 shows the results of these measurements. The median difference from the target was 0.68 degrees (standard deviation 2.41 degrees) in azimuth and 0.61 degrees (standard deviation 3.24 degrees) in elevation. According to this, the direction indication method is approximately as accurate as the human auditory localization accuracy as its best.

5.1.2 Static sound source

In the static experiment[P5], there were 14 real source (loudspeakers) positions, and 14 virtual source positions between loudspeakers. Each position was presented twice with the screen compensation on and twice without it for each subject. In total, each subject had to localize 112 source positions. In this experiment, a pink noise burst was used as an auditory stimulus, and no visual stimulus were used.

Angular position of each presented sound source and the corresponding pointing response were recorded. For analysis, and for the Figures 5.3, and 5.4, each measurement was rotated so that the presented source was at the coordinate system origin (in front of the subject). This was motivated by the fact that the subjects were instructed to turn towards the sound source. Using this coordinate system, we calculated the error angles as deviations from the source azimuth and elevation or in other words origin.

The median of the error angles for real sources without screen compensation was 9.5 degrees and with screen compensation it was 8.5 degrees. Median of error angles for virtual sources was 16.3 degrees and 15.5 degrees with and without screen compensation, respectively. Figure 5.2 shows more detailed statistics with the error angle broken down into its azimuth and elevation components. All medians of errors for real and virtual sources with and without screen compensation are presented in Table 5.1.

Table 5.1: *Median of error (in degrees) for real and virtual static sound sources with and without screen compensation. [P5]*

	Azimuth	**Elevation**	**Error angle**
with screen compensation			
Real source	4.5^o	5.7^o	8.5^o
Virtual source	4.9^o	11.6^o	15.5^o
without screen compensation			
Real source	4.4^o	7.3^o	9.5^o
Virtual source	6.5^o	9.5^o	16.3^o

For the virtual sources the azimuth accuracy is slightly better with screen compensation, however, the elevation accuracy is slightly better without screen

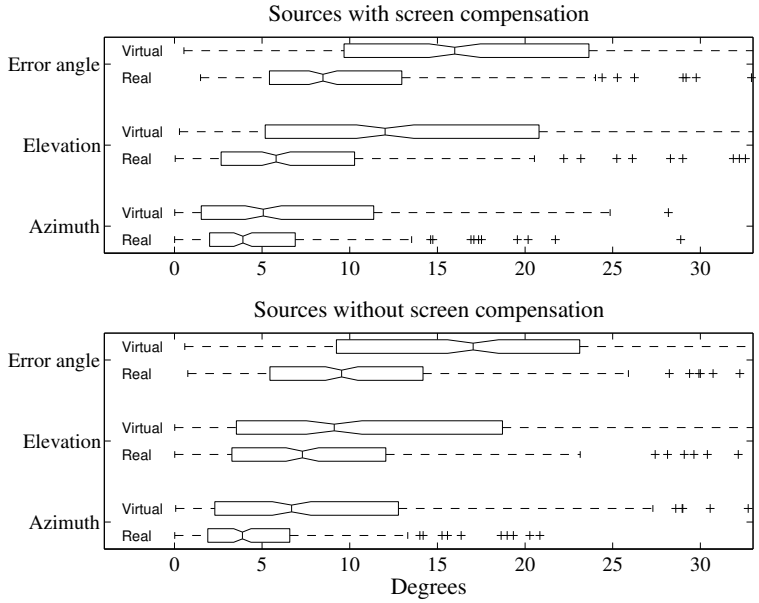

Figure 5.2: *Box plots of pointing error for real and virtual static sources with and without screen compensation [P5].*

compensation. These differences, however, are smaller than the deviation of our direction indication method. For both source types, the measured response errors with and without screen compensation are in the same area and the deviations are similar to those seen in Figures 5.3 and 5.4. The similarity of accuracy and deviations is clearly seen in Figure 5.2.

We analyzed three factors from the measured data: effect of screen compensation, source type, and subjects. According to Lilliefors test [61], this data set was not normally distributed, and therefore, it does not fulfill the analysis of variance (ANOVA) assumptions. Thus, we applied the Kruskal-Wallis test, which is a non-parametric version of the one-way ANOVA. Exact p-values obtained in the analysis with corresponding H0:hypotheses are presented in Table 5.2.

Localization accuracy for the real sources and the virtual sources is statistically significantly different for azimuth error ($\chi^2 = 24.62$, $p < 0.05$), elevation error ($\chi^2 = 36.18$, $p < 0.05$), and error angle ($\chi^2 = 96.91$, $p < 0.05$).

In the statistical analysis, the effect of screen compensation is not significant for azimuth error ($\chi^2 = 1.73$, $p > 0.05$), elevation error ($\chi^2 = 0.42$, $p > 0.05$), and error angle ($\chi^2 = 0.70$, $p > 0.05$). Since the screen compensation did not affect localization accuracy, we kept it on in the follow up experiment with moving sound sources, because it equalize the frequency response.

Differences between the subjects are statistically significant for azimuth

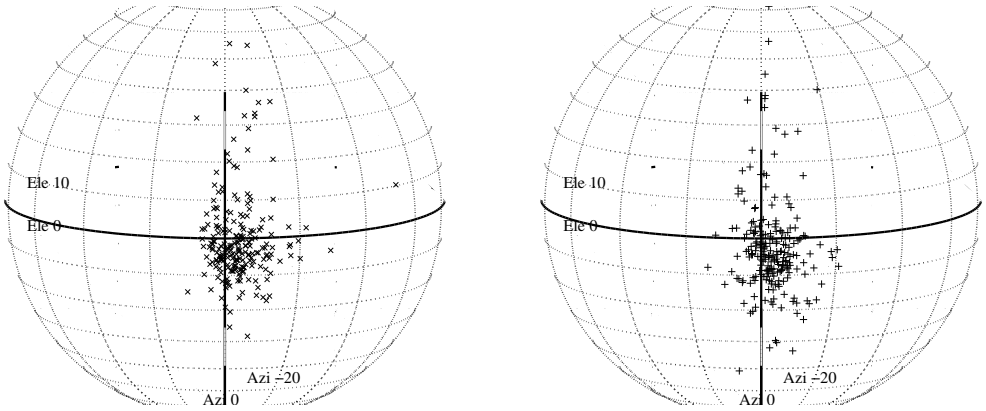

Figure 5.3: *Measured response errors for real sources (loudspeakers). The resolution of the azimuth grid is 20 degrees and elevation grid 10 degrees. On the left values with screen compensation 'x' and on the right values without screen compensation '+' [P5].*

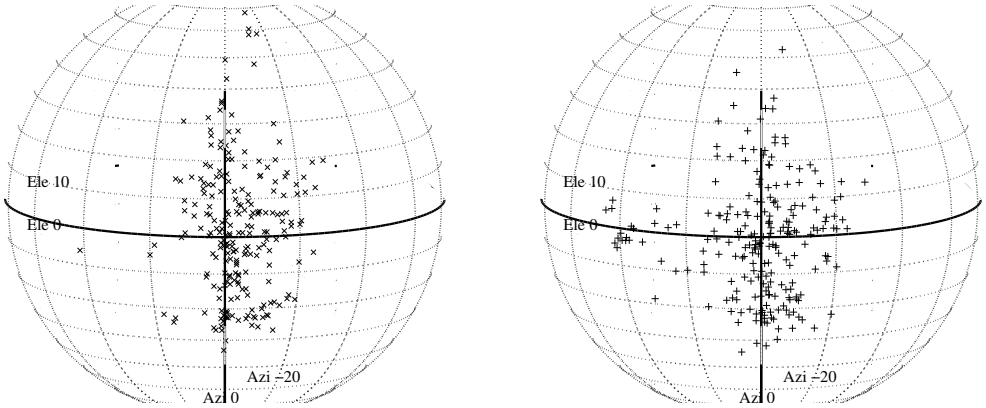

Figure 5.4: *Measured response errors for virtual sources. On the left values with screen compensation and on the right values without screen compensation [P5].*

error ($\chi^2 = 44.07,\ p < 0.05$), elevation error ($\chi^2 = 33.10,\ p < 0.05$), and error angle ($\chi^2 = 46.06,\ p < 0.05$).

5.1.3 Moving sound source

In the moving sound source experiment[P5], two different trajectories (see Figure 5.10) and three different auditory stimuli were used. Each path-stimulus combination was presented four times to each subject. Each subject conducted 24 trajectory tasks.

The error values for successive samples in a trajectory are dependent on each other, because of physical inertia. Statistical analysis methods assume independent values. To achieve this, we took median values for each separate trajectory and used these values.

Table 5.2: *H0 hypotheses and p-values obtained with Kruskal-Wallis test. Statistically significant p-values less than 0.05 are in bold [P5].*

	Azimuth	**Elevation**	**Error angle**
H0: Accuracies for real and virtual sources are not different	**0.00**	**0.00**	**0.00**
H0: Screen compensation does not affect on accuracy			
All sources	0.19	0.52	0.40
Real source	0.69	0.13	0.36
Virtual source	**0.03**	**0.03**	0.65
H0: Subjects do not differ in accuracy			
All sources	**0.00**	**0.00**	**0.00**
Real source	**0.00**	**0.00**	**0.00**
Virtual source	**0.00**	**0.03**	0.16

Table 5.3: *Median error in degrees in the moving sound source experiment for each stimulus and in general [P5].*

	Azimuth	**Elevation**	**Error angle**
Noise	12.8^o	13.2^o	21.8^o
Music	14.6^o	11.7^o	21.7^o
Frog	12.6^o	10.1^o	19.3^o
All	13.4^o	11.5^o	20.8^o

The overall medians and the medians for each separate stimulus are presented in Table 5.3. The overall median of error was 13.4 degrees in azimuth, 11.5 degrees in elevation, and 20.8 degrees in error angle. Figure 5.5 shows more detailed statistics with the error angle broken down to its azimuth and elevation components.

We analyzed two factors from the measured data: effect of signals and subjects. As in the static case, the data set was not normally distributed, and therefore we applied the Kruskal-Wallis test. Exact p-values obtained in the analysis and corresponding H0 hypotheses are presented in Table 5.4.

Differences in measured localization accuracy between the subjects are statistically significant for azimuth error ($\chi^2 = 15.33,\ p < 0.05$), elevation error ($\chi^2 = 81.13,\ p < 0.05$), and error angle ($\chi^2 = 28.05,\ p < 0.05$). Differences between the stimuli are also statistically significant for azimuth error ($\chi^2 = 7.79,\ p < 0.05$), elevation error ($\chi^2 = 9.90,\ p < 0.05$), and error angle

($\chi^2 = 7.95$, $p < 0.05$).

Table 5.4: *H0 hypotheses and exact p-values obtained with Kruskal-Wallis test [P5].*

Azimuth	Elevation	Error angle
H0: There is a no difference between the subjects		
0.03	**0.00**	**0.00**
H0: There is a no difference between the stimuli		
0.02	**0.01**	**0.02**

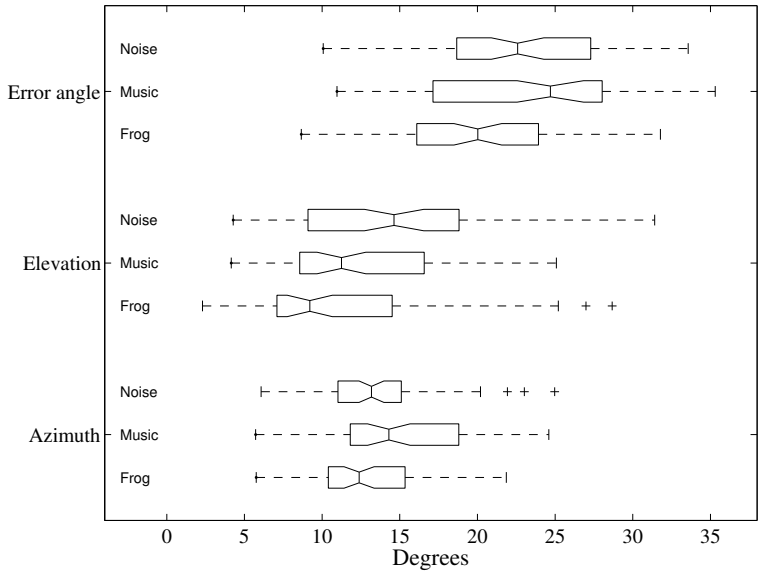

Figure 5.5: *Box plots of error in the moving sound source experiment for each stimuli [P5].*

As seen in Figure 5.6, the achieved accuracy in azimuth is more or less the same for each subject. However, the most accurate subject in elevation has a median error (6.6 degrees) that is almost three times smaller than the error of the most inaccurate subject (18.6 degrees). The most inaccurate subject in elevation was the most accurate subject in azimuth.

Considering the dynamic behavior of pointing, there is a latency in start and turns as seen in Figure 5.10. For example in the time-azimuth and time-elevation plots on the left, there is a three to four seconds delay, before the subjects start to follow the movement of the target source. In both time-azimuth plots, it is possible to see that responses follow the moving source with delay. When the azimuth is increasing, most of the measured values are

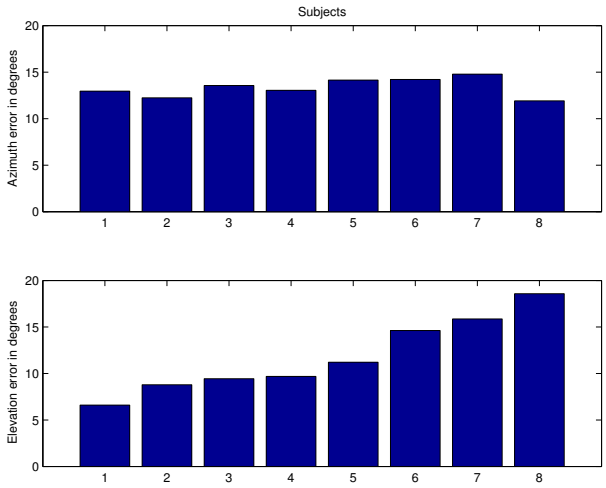

Figure 5.6: *Azimuth and elevation error medians in the moving sound source experiment for each subject. Subjects are ordered according to their elevation accuracy [P5].*

smaller than the source azimuth value, and when the azimuth is decreasing the measured values are greater than the source value. This is not clearly seen in time-elevation plots.

In Figure 5.10 on the left, the measured trajectories are bent at their latter ends towards a loudspeaker located -90 degrees in azimuth and -39 in elevation. In the time-azimuth plot, the average measured end azimuth value varies from -90 to -100 degrees. The end azimuth value of the target sound was -110 degrees.

On the right of the time-azimuth plot, there is a stepwise behavior in the trajectory. In addition, the movement of measured location continued after the source stopped.

Comparing the time-azimuth and time-elevation plots, it is easy to see that subjects have less variability in their azimuth accuracy than in their elevation accuracy. The box plots in Figure 5.5 support this.

5.1.4 Moving sound source effect of the distracting auditory stimulus

In the distracting auditory stimulus experiment [P4], there were three different target trajectories, one static distracting sound position and two different distracting sound trajectories and three different stimuli. For each subject, each possible combination was presented once, which equals 27 tasks per subject.

In [P4], the effect of the distracting auditory stimulus in localization of the moving sound source was measured. In comparison with results achieved in [P3], the median azimuth error increased almost five degrees (from 12.5

to 17.0). In the median elevation error, the increase was insignificant (from 24.1 to 25.4). Also, in this experiment the differences between the signals were insignificant.

5.2 Navigation

The main results of the audio-visual, and different auditory cue navigation experiments are described in this section. These experiments are described in more detail in publications [P6, P7].

For each subject, the number of gates found during the run was recorded. In addition, the time and location information were recorded. From the measured data three factors were analyzed: the number of found gates, search time between the gates, and normalized path length[2]. The statistical analysis was not made with analysis of variance (ANOVA), since group variances were not equal. The Kruskal-Wallis test was applied instead.

5.2.1 Comparison of audio-visual, visual and auditory navigation [P6]

Nine non-paid subjects completed the test so that each of the three cues was presented twice to each subject. The results shows that navigation with audio-visual cues is remarkably easier and faster than with auditory or visual cue alone. Median values of found gates in three minutes were 8 (auditory cue), 18.5 (visual cue), and 28 (audio-visual cue) gates. This difference is statistically significant ($p = 0.00$) and all three group means differ from each other. The same statistically significant difference was also found for the search times and normalized path lengths. These differences are seen in box plots in Figure 5.7.

5.2.2 Comparison of different auditory cues [P7]

Four different auditory stimuli (*Gain, Gain+Pitch, Gain+Rate,* and *All*, see section 4.3.2) were applied in this experiment. Each cue type was presented twice to each subject (N = 8) in randomized order. The statistical analysis was performed with the same methods as in audio-visual test.

Although with the *Gain+Pitch* stimulus the number of found gates was the largest, the difference is not statistically significant ($p = 0.10$). Difference in search times and normalized path lengths are statistically significant ($p = 0.00$ and $p = 0.00$). Post hoc analyzes showed that with search times the mean ranks of *Gain+Pitch* and *All* differ significantly from the mean rank of *Gain*. Correspondingly, the mean rank of *Gain+Pitch* differs significantly from the mean ranks of *Gain* and *Gain+Rate* with normalized path lengths. Box plots of these are seen in Figure 5.8.

[2]normalized path length = travelled path length / distance between gates

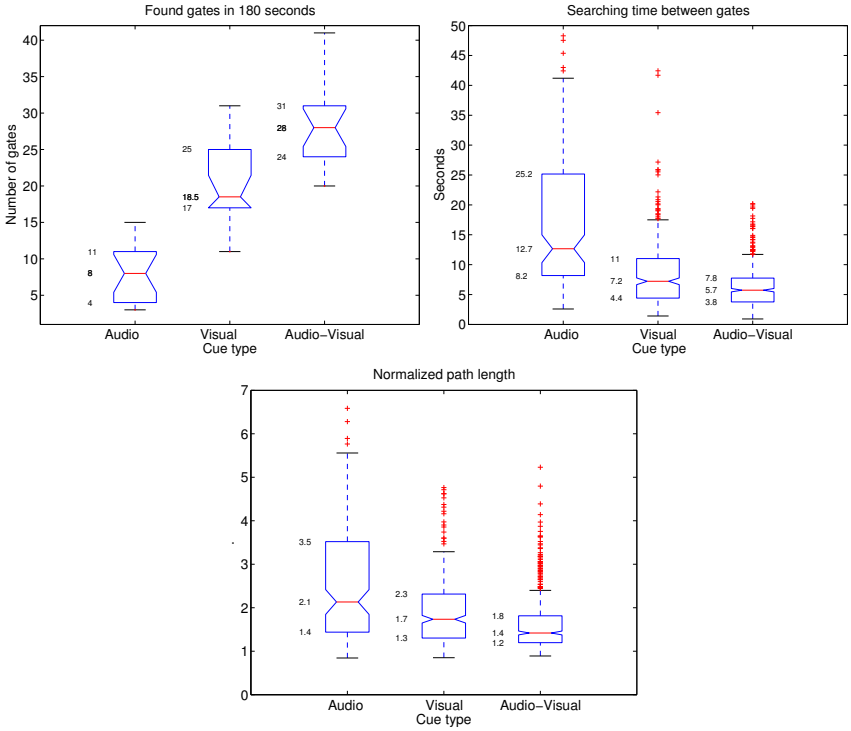

Figure 5.7: *Results of the audio-visual navigation test [P6].*

5.3 Orientation

In the orientation experiment [P8], there were four types of conditions: visual, gain, rate, and pitch. In the visual condition, a subject went through the route withour any auditory cues. The experiment consisted of two test sets and each condition was used once for each subject in both sets. Experiment details are described in [P8].

The amount of disorientation is measured and analyzed using absolute values of recorded pitch and roll angles (defined in Figure 2.1). In the analysis, the medians of the absolute values of angles throughout the route were used.

The main results are from the analysis of the second set. The amount of time to accomplish the task is not condition dependent ($p = 0.73$). With pitch angle, the difference between the visual condition and auditory conditions is statistically significant ($p = 0.00$). With roll angle, the difference between the visual condition and auditory conditions is smaller than with pitch angle error, but the difference is statistically significant ($p = 0.04$). Differences between the visual condition and auditory conditions are seen in Figure 5.9.

There were a big differences between the subjects. With roll angle error, the most accurate subject was more accurate with his worst condition than

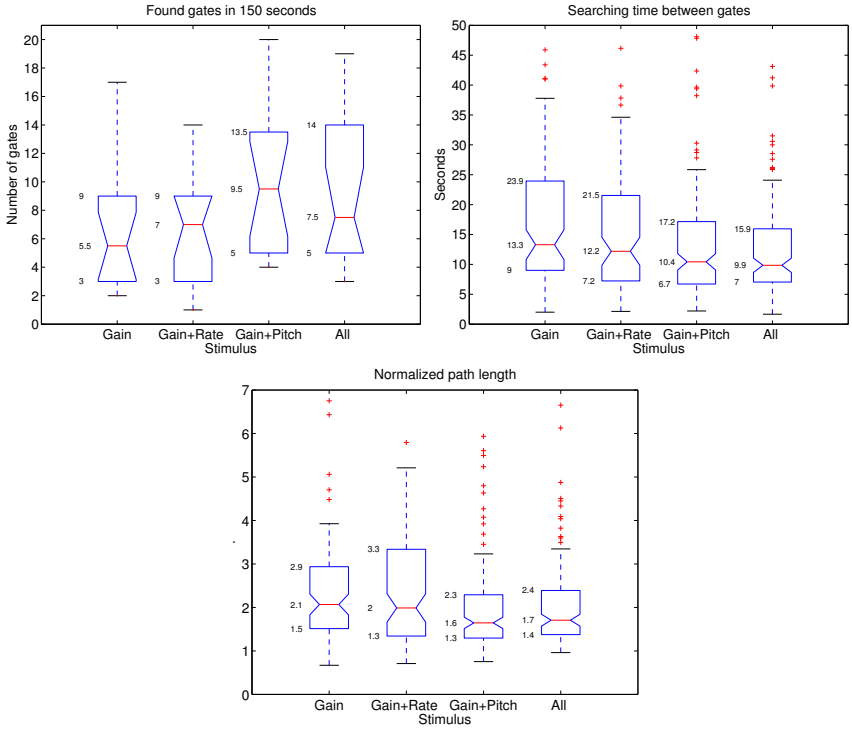

Figure 5.8: *Results of the different auditory stimulus signals. [P7]*

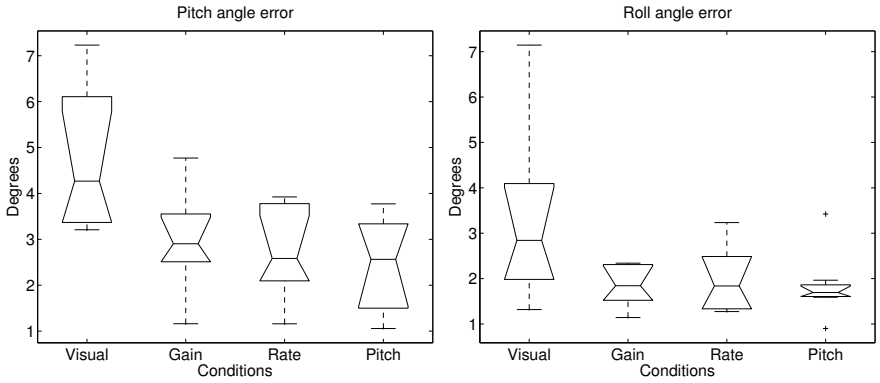

Figure 5.9: *Box plots of absolute value pitch and roll angle error for each condition in the second test set. [P8]*

the least accurate subjects with their most accurate conditions. In particular, there was a lot of variation in the visual condition.

After the test, subjects were asked to put the auditory conditions in subjective order. In this evaluation, all eight subjects put conditions in the same order: gain (best), pitch, and rate.

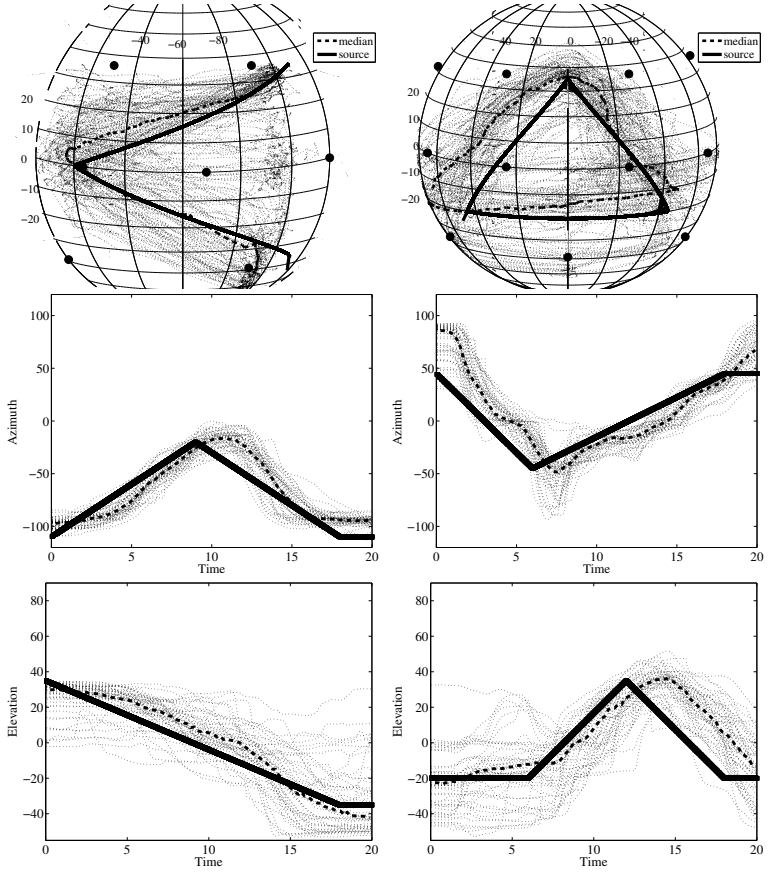

Figure 5.10: *Plots of trajectories used in the moving sound source experiment. The position of the simulated sound source is indicated with a thick black line and the measured trajectories are indicated with dotted lines. Medians of the measured trajectories are indicated with a thick dashed line. In azimuth-elevation plots, the locations of the loudspeakers are indicated with black circles. The resolution of the azimuth grid is 20 degrees and elevation grid 10 degrees. Waypoints of trajectories are defined as follows* **Left column:** *the starting point (-110 degrees in azimuth, 35 degrees in elevation), the turning point (-20, 0) and the end point (-110, -35).* **Right column:** *the starting point (45, -20), the first turning point (-45, -20), the second turning point (0, 35) and the end point of the target sound is the same as the starting point. [P5]*

6 Discussion

6.1 Localization

In this section, the analysis of results of the static sound source [P5], moving sound source [P5], and moving sound source with distracting stimulus [P4] experiments is provided. Afterwards, a discussion about the other issues is included.

6.1.1 Static sound source

We compared our results with previous results achieved using pointing as a direction indication method. As seen in Table 6.1, localization accuracy of static real sources in EVE was in line with the accuracy that Sandvad [34] achieved with real sources in an anechoic chamber. According to Pulkki [15] this is expected, because results achieved in anechoic chambers can be qualitatively applied in moderately reverberant conditions like a virtual room due to the precedence effect.

In the static experiment [P5], the median localization error is smaller with real sources than with virtual sources as expected, because virtual sound source is distributed to three loudspeakers and therefore it is spatially blurred. In addition, measured localization accuracy with static virtual sources was in line with the accuracy achieved with individualized HRTFs by Sandvad [34] and Djelani et al. [32]. This was better than expected, but understandable due to the fact that each listener uses his/her own ears in loudspeaker reproduction. Martin et al. [35] reported localization accuracy of the virtual sources that is comparable with the localization accuracy of the real source. As expected, localization accuracy in experiments reported in [P5] was smaller than accuracy in optimal conditions.

In experiments [P3-P8], there were no front-back confusions or other errors generated by the cone of confusion. These were avoided because subjects could freely move and rotate their head during the experiments.

In general, screen compensation did not affect localization accuracy. The only exception was the situation where the virtual source was reproduced using one visible loudspeaker and two others behind the screens. In that case and without the screen compensation, the timbre difference between the three used signal sources was too large to produce a single virtual source due

to the stream segregation effect [62]. Subjects reported that instead of one stimulus, they heard two simultaneous stimuli. With screen compensation, subjects localize this source as accurately as other virtual sources.

In real source plots in Figure 5.3, the centers of the measured locations are slightly shifted to the right and down. The shift to the right suggests that although our direction indication device was equipped with a handle and sight, the right-handed people (all our subjects) pointed slightly away from the real location. Future research is needed to find out, how to design a direction indication method that is insensitive to handedness.

Table 6.1: *Results (in degrees) from different localization experiments using pointing as a direction indication method.*

	Azimuth	**Elevation**	**Error angle**
Static experiments			
Sandvad 1996			
Real source	7^o	9^o	
HRTFs	7^o	10^o	
Djelani et al. 2000			
HRTFs			12.4^o - 21.1^o
Martin et al. 2001			
Real source			8.0^o - 11.0^o
HRTFs			9.6^o - 9.7^o
Our experiment			
Real source	4.5^o	5.7^o	8.5^o
Virtual source	4.9^o	11.6^o	15.5^o
Moving sound source experiment			
Our experiment	13.4^o	11.5^o	20.8^o

There was a statistically significant difference in accuracy between the subjects. This is due to the fact that each person's ears, head, and torso are individual, and, therefore, each person receives an individual set of directional cues. In addition, this is in line with results achieved in other experiments. Pulkki has found in his loudspeaker listening tests [37], that especially the accuracy of elevation perception is highly individual. In headphone experiments with HRTFs [23], the similar variations between subjects were reported.

6.1.2 Moving sound source

In comparison with static virtual sources, the increase in median error angle is comparable with the amount of minimum audible movement angle [18]. In the moving sound source experiment [P5], the achieved elevation accuracy was the same as that of static virtual sources (Table 6.1). The achieved

azimuth accuracy was decreased compared to the static experiment. One explaining factor is that in all experiments subjects were allowed freely turn in the middle of the EVE and therefore. in the static experiment subjects could optimize their listening orientation for each position. With a moving sound source this is not possible except for starting positions.

Although the measured azimuth error is slightly larger than the elevation error, subjects showed less variation in their azimuth accuracy than in their elevation accuracy. This is shown in Figure 5.5; the boxes for the azimuth accuracy are smaller than the boxes for elevation accuracy. The same difference in variation applies also to the static experiment as shown in figure 5.2. Box-plots do not reveal the time dependent features of the moving sound sources. For further analysis, it is better to compare the time-azimuth and time-elevation plots in Figure 5.10. Also in these plots, the subjects have less variation (i.e., measured trajectories are closer to each other) in their azimuth accuracy than in their elevation accuracy.

The difference between the localization accuracy of the static and moving sources can also be explained by other factors: the reaction delay at the beginning of the trajectory and in the slow reactions to changes in the direction of the trajectory, as is evident in Figure 5.10. In both time-azimuth plots the responses follow the moving source with delay. When the azimuth is increasing, most of the measured values are smaller than the source azimuth value, and when the azimuth is decreasing, the measured values are greater than the source value. This lag decreases the localization accuracy systematically. Figure 5.10 also suggests that direction changes in the azimuth are better perceived than in the elevation. Some part of the latency is explained by the normal latency in human reaction time. For further analysis, similar experiments should be accomplished with moving visual targets. This is one area for future research.

The difference between the stimuli was statistically significant (0.05 confidence level) but not highly significant (0.01 confidence) for the azimuth error, or for the error angle. This is also shown in Figure 5.5. The notches in the box plots overlap especially in the azimuth and error angle cases. In addition, the difference between the subjects in azimuth error was not statistically highly significant, as shown in Figure 5.6. Individual azimuth error medians are close to each other while there is a large difference in elevation error medians. Pink noise was not the most accurate stimulus, as expected. The frog stimulus seems to provide the most accurate results. One explanation is that the subjects reported the frog stimulus to be the most irritating stimulus of all. When a person gets irritated his/her attention level increases, and this increases also the performance.

Trajectories measured in the moving sound source experiment suggest that loudspeaker positions have an effect on trajectories. Measured trajectories have a tendency to bend toward the loudspeaker positions. In Figure

5.10 in the left column, the ends of the measured trajectories are bent towards a loudspeaker located in position -90 in azimuth and -39 in elevation. In the time-azimuth plot, the average measured end azimuth value is about -95 degrees. In contrast, the end azimuth value of the target sound was -110 degrees.

This bending may be partly explained by the masking effect [1] (page 223). In the research environment, there is a background noise due to the air conditioning. When the virtual source location is near one of the loudspeakers, the signals from the other two loudspeakers could be masked by the background noise, and the subject could hear only the signal from the one loudspeaker.

6.1.3 Moving sound source with distracting auditory stimulus

With interfering noise, the localization blur is dependent on signal levels and frequencies [1]. If the level of the target signal is about 10 - 15 dB above the interfering noise, the localization blur is at the same level as it would be without the interfering noise. On the other hand, in an article by Tuyen and Letowski [63] it was mentioned that a 6 dB signal-to-noise ratio is appropriate for tasks requiring accurate frontal localization. In the experiment described in [P4], the distracting stimulus increased the localization blur, although the gain difference between the target sound and distracting sound was 10 dB. In most cases the distracting stimulus only decreased the accuracy of localization. In ten percent of the cases there was confusion when the subject temporarily pointed to the distracting stimulus instead of the target.

In [P4], the same stimulus was used as a target sound and as a distracting stimulus. Subjects were expected to have more confusions with sounds. In the experiment, the target sound was played before the distracting sound. As a result, subjects concentrated on the target sound and the effect of the distracting sound was diminished.

Distracting auditory stimulus results were measured before the changes in research environment. In Table 6.2, the comparison of results measured in the moving sound source experiment before the changes in research environment [P3] and after the changes [P5] are shown. The median of the elevation error decreased from 24.1 to 11.5 degrees, but the standard deviation is almost the same. In azimuth the differences are insignificant. Since the standard deviation of the measured values does not change remarkably, the results achieved in the distracting auditory stimulus experiment can be considered to be reliable.

6.1.4 Other issues

In this section the selection of the signal, reproduction, perception and measurement method is discussed.

Table 6.2: *Changes in measured results after changes in a design of the virtual room.*

	Azimuth		**Elevation**	
	P3	P5	P3	P5
Median of amount of error	12.5^o	13.4^o	24.1^o	11.5^o
Standard deviation	18.4^o	19.1^o	19.4^o	16.2^o

In an informal evaluation of the virtual room (section 4.2.1) different kinds of signals (high- and low-pass filtered noise and some filtered speech signals) were used. For example, signals having frequencies only above 3 kHz were perceived all the time above horizontal plane, even when they are reproduced using loudspeaker in the horizontal plane or below the horizontal plane. This in line with findings by Middlebrooks [17]. In his article, he mentioned that manipulations of the sound source spectrum can generate erroneous localization responses especially in elevation.

According to experience achieved during this research, designing and evaluating the virtual room is crucial to achieve a good localization accuracy. Before enhancements described in section 4.2.1, the median of amount of elevation error varied from 12.1 degrees (static real source) up to 24.1 degrees (moving sound source). After enhancements the median of the amount of elevation error varied from 5.7 degrees (static real source) up to 11.6 degrees (static virtual source). Enhancements did not affect measured azimuth accuracy.

Another thing to consider is the effect of the loudspeaker configuration in a virtual room. The problem in a virtual room is that due to a visual display configuration, there are many areas where one cannot set loudspeakers. Ballas et al. [64] explored the effect of auditory rendering on perceived movement. According to their experiment, increasing the number of loudspeakers in VBAP reproduction enhanced the accuracy in perceived movement. In EVE it might be possible to add additional loudspeakers to achieve better accuracy, but it is a topic for future experiments.

The proper direction indication method was found through an iterative process. In the preliminary experiment in [65], a wand was used for pointing. It was found that a wand is not suitable to use as a direction indication device. In following experiments [P2-P4], a tracked baton was used and results were more accurate. For the final experiments [P5], the handle and sight were added to the tracked baton, and the most accurate results were achieved with this direction indication method.

Other researchers have shown that pointing [32, 35] is more accurate than other direction indication methods. This suggests that direction indication method affects results and a poorly chosen direction indication method

will decrease the reliability of the localization results.

Pulkki [37] found in his listening tests that perceiving elevation of a virtual source varies between the subjects. Results in this thesis support his findings. There were remarkable differences between the subjects. The most accurate subject had a median error that was much less than the median error of the most inaccurate subjects. Even though there was a large variation between the subjects, each subject was consistent and in the moving sound source experiment the same trajectory for all stimuli was perceived for each subject. Impact of the subject's level of experience was not measured. It is one possible area for the future research.

6.2 Navigation

The navigation paths in the audio-visual experiment [P6] suggested that the subjects used the auditory cue (if available) to define the approximate direction to the target gate, and the visual cue (if available) in the final approach. This is natural because auditory cue can be used when the target is not visible which may happen at first. Visual cue is more accurate and therefore is used in the final approach. When both cues were available navigation from one gate to the next was a straightforward task.

The navigation paths in the auditory cue experiment [P7]with *Pitch* information are less complex than without it. In particular, the height changes smoothly when approaching the target gate.

The results of the auditory cue experiment [P7] in section 5.2.2 prove that the assumption about the difficulty of hearing the height of a sound source was correct. When elevation information is encoded with the navigation cue signal, subjects found gates faster and length of the travel paths were shorter.

When the results of the audio-visual (section 5.2.1) [P6] and auditory cue [P7] experiments are compared an interesting result can be found. The normalized travel path lengths with *Visual* cue are close to *Pitch* and *All* stimuli results (no significant difference, $p = 0.50$). However, search times were longer with auditory navigation (highly significant difference $p = 0.00$). In any case, our results suggested that auditory navigation is almost as easy as visual navigation.

The choice to use noise bursts as a base auditory stimulus was based on results from our preliminary navigation experiment [50] and was supported by results from Walker et al. [51]. In their follow up experiment, Walker et al [52] explored the effect of the capture radius. In experiments described in [P6, P7], a fixed size capture radius was used in both experiments. The size of the radius was chosen during informal tests before the reported experiments. In this informal test, it was found that the size of the radius is a trade-off between the speed and accuracy. Results by Walker et al., confirm these informal findings.

6.3 Orientation

All the subjects understood the auditory artificial horizon [P8] immediately and found it intuitive to use. Most of the subjects needed only one training round for each auditory condition before they started the experiment test sets. No significant difference was found in performance times under different conditions. This suggests, that auditory cues did not increase the cognitive load of the subjects.

The amount of disorientation was larger in pitch angle than in roll angle for each condition. The difference varies from 0.7 degrees (gain condition) up to 1.6 degrees (visual condition). This difference suggest, that subjects used the horizontal visual cues in front of them to keep the roll angle oriented. The pitch angle was harder to keep oriented, especially when subjects were moving up or down the aisles.

Orientation accuracy did not change between the two test sets in visual condition. In auditory conditions, subjects performed better in the second test set than in the first test set (see for example table 3 in [P8]). This suggests that subjects had learned to use the auditory artificial horizon during the experiment.

In subjective ranking, the gain condition was preferred. Although pitch condition was subjectively ranked second best, two of the subjects reported that it was annoying. Rate condition was found least useful because it did not contain a clear reference value for the perfect orientation as did other two auditory conditions. This result suggests, that providing reference value information, is an important part of designing auditory stimulus. This supports the point made by Walker and Kramer [66] that it is crucial to test any sound design and not rely on the intuitions.

The use of auditory artificial horizons in real usage scenarios is an open question. One can argue that it will disturb the communication if there are group of people exploring the model.

7 Summary and conclusions

This work is one of the first studies to explore the localization accuracy of loudspeaker reproduced sounds in a virtual room. In addition, spatialized auditory cues in navigation have been successfully applied. Finally, a working auditory artificial horizon was designed and tested. In this chapter the summary of the main results, a few potential applications, and directions for the future research are presented.

7.1 Main results

The basic research topic was to measure the localization accuracy in a virtual room. The main results of the localization experiments can be summarized as follows:

- Localization accuracy of static real sources in a virtual room is as good as the localization accuracy of static real sources in an anechoic chamber.
- Localization accuracy of loudspeaker reproduced static virtual sources is as good as the localization accuracy of static virtual sources reproduced using individualized HRTFs.
- Localization accuracy of the moving sound source is less accurate than the localization accuracy of the static virtual source. The difference is comparable to the minimum audible movement angle.
- Additional distracting auditory stimulus decrease the localization accuracy of the moving sound source.

Main results of the navigation experiments are:

- Auditory navigation is possible in a 3D environment even without any visual cues.
- Simultaneous visual and auditory cues support each other in navigation and are more efficient than navigation using visual or auditory cues alone.

- Providing additional auditory coded distance and elevation information increases the subject's auditory navigation performance.

The main result of the orientation experiment is:

- The auditory artificial horizon helps subjects keep the model better oriented than without it.

In addition to scientific results, the following findings were made during the experiments:

- Direction indication method affects measured localization accuracy.
- Design of the virtual room affects localization accuracy. Design includes reflection minimization and finding a proper calibration signal.
- Screen compensation does not affect localization accuracy, but it improves perceived quality.
- Subjects prefer auditory cues that clearly indicate important reference values. For example, in the orientation experiment, two of the auditory cues provided accurate information, when the model was in the correct position. Subjects rated these two better than the third one that did not provide the reference value.

7.2 Applying the results

Although the achieved localization accuracy of virtual sources was not at the level of the human perception, it is accurate enough to enable auditory navigation and use of auditory artificial horizons. This suggests that spatial sound reproduction could be used more often than currently in different virtual reality application areas mentioned in section 1. In flight simulators, it can be applied to inform the pilot about other planes and their location. In virtual environment games, the player could localize targets and other important objects according to their directional auditory cues. In scientific visualization, point-like auditory beacons could be used, for example, to 'highlight' interesting parts of the models, such as important molecular structures in large molecular complexes or critical pressure values in complex 3D computational fluid dynamics data sets. As with the virtual environment games, spatialized auditory cues help the user to localize the most interesting molecules, component and targets, which gives the user better insight to data.

Auditory artificial horizon could be used in tasks where the upright position is crucial to the task and visual feedback is not available and/or feasible. One potential group of users for an auditory artificial horizon and auditory navigation systems is visually impaired people.

Since the achieved localization accuracy was at the same level as with individualized HRTFs in headphone reproduction, it can be assumed that auditory navigation and an auditory artificial horizon should work in virtual environments applying headphone reproduction. Further research is needed to support this.

Localization accuracies of different virtual rooms and audio reproduction systems could be estimated in a short period of time using the signals and interactive method described in section 4.2.1.

7.3 Future directions

There are three main directions for future research. The first is applying auditory navigation and an auditory artificial horizon in different kinds of virtual environments. The second is to employ spatial audio in other common tasks in virtual reality applications and the third is to explore in more detail multi-stimuli and multimodal situations.

All the experiments in this research were accomplished in a single virtual room. Further research is needed to determine if the auditory navigation and auditory artificial horizon can be applied in other kinds of virtual environments with different kinds of spatial audio reproductions methods, visual displays, and interaction devices and methods.

In this research, localization, navigation and orientation were explored. As described in Chapter 1, there are other common tasks in virtual reality application like data representation, object selection, and manipulation, in which the spatial audio could be employed.

Except the distracting auditory stimulus localization experiment, all the experiments were accomplished using one auditory stimulus at a time. Further experiments will be needed to show how well subjects could perform in multi-stimuli tasks. In navigation experiment effect of combined audio-visual stimuli was explored. In this experiment, auditory and visual cue supported each other. Further research is needed to determine how the subjects will perform in a situation of conflicting audio-visual stimuli. Another even larger open research area is to explore the interaction of combined visual, auditory, and tactile stimuli.

Summary of publications and author's contribution

This chapter summarizes the publications incorporated in this thesis and describes the author's contribution. The author of this thesis is the primary author of all the publications, with the exception of [P7]. The author is the sole author of publications [P2] and [P4].

Publication [P1]
M. Gröhn, T. Lokki, L. Savioja, and T. Takala.
Some Aspects of Role of Audio in Immersive Visualization.
In *Visual Data Exploration and Analysis VIII, Proceedings of SPIE Vol. 4302*, pages 13–22, San Jose, Jan. 2001.

This article describes the field and basic concepts of this research. The four common tasks in immersive visualization: localization, navigation, orientation, and sonification are defined in this article. Spatial sound reproduction can be applied in each of these tasks.

The author is responsible for sections 1-4, and section 6 of this article. Section 5 summarize the preliminary navigation experiment that was accomplished together with other authors and is originally published in [50].

Publication [P2]
M. Gröhn.
Is Audio Useful in Immersive Visualization?
In *Stereoscopic Displays and Virtual Reality Systems IX, Proceedings of SPIE Vol. 4660*, pages 411–421, San Jose, Jan. 2002.

This article examines the localization results of the static sound sources in a virtual room using a tracked baton as a pointing device. The median absolute azimuth error was 6.9 degrees, and elevation error was 15.9. degrees.

The author was the sole contributor and researcher in this study except for sections 5.1 and 5.2. In these sections there is a short summary of the preliminary localization experiment published in more detail in [65].

Publication [P3]

M. Gröhn, T. Lokki, and T. Takala.
Static and dynamic sound source localization in a virtual room.
In *Proceedings of AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, pages 337–344, June 2002, Espoo, Finland.

The main outcome of this article is the localization results of the moving sound sources. The median absolute azimuth error for a moving source was 12.5 degrees, while elevation error was 23.4. degrees. Localization accuracy of starting points was similar to the localization accuracy of static sources in [P2].

The author was responsible for this experiment and performed the majority of the work. Tapio Lokki and Tapio Takala assisted in test design.

Publication [P4]

M. Gröhn.
Localization of a moving virtual sound source in a virtual room, the effect of a distracting auditory stimulus.
In *Proceedings of the International Conference on Auditory Display (ICAD 2002)*, pages 394–402, Kyoto, Japan, 2.-5. Jul. 2002.

The findings in this article include the effect of a distracting auditory stimulus in the localization results of the moving sound sources. With a distracting stimulus the median absolute error for a moving source was 17.0 degrees, while elevation error was 25.4 degrees. As expected distracting auditory stimulus decreased the accuracy compared to results achieved in [P3].

The author was the sole contributor and researcher in this study.

Publication [P5]

M. Gröhn, T. Lokki, and T. Takala.
Localizing loudspeaker reproduced sounds in a cave-like room.
Accepted to published in *Presence: Teleoperators and Virtual Environments, Vol. 16*, 2007.

The main result of this article was the localization results after the changes in a test environment, and the direction indication method explained more in detail in section 4.2.1. In the localization results reported in papers [P2-4], there were systematic errors in elevation. Perceived locations were above the location of the sources. Before enhancements the median of the absolute error for the azimuth was 6.9 degrees and for elevation 15.9 degrees. After enhancements the value for the azimuth was 4.8 degrees, and for elevation 10.8 degrees. With a moving source, prior to the enhancements the value for

the azimuth was 12.5 degrees and for elevation 24.1 degrees, while after the enhancements the value for the azimuth was 13.4 degrees, and for elevation 10.0 degrees. An accurate direction indication method was developed as a side product of this experiment.

The author was responsible for this experiment and performed the majority of the work. Tapio Lokki and Tapio Takala assisted in test design.

Publication [P6]

M. Gröhn, T. Lokki, and T. Takala.
Comparison of auditory, visual and audio-visual navigation in a 3D space.
In *Proceedings of the International Conference on Auditory Display ICAD 2003*, pages 200–203, Boston, 6.-9. Jul. 2003.
Reprinted in ACM Transactions on Applied Perception Vol. 2 Nr. 4, 2005.

This article compared auditory, visual and audio-visual navigation in a 3D space. Audio-visual navigation was clearly most efficient. Visual navigation was second, and the auditory navigation was least efficient. Further analysis of the travel paths indicated that the auditory cue was used in the beginning to locate the next target, and the visual cue was the most important in the final approach to the gate.

The author was responsible for this experiment and performed the majority of the work. Tapio Lokki and Tapio Takala assisted in test design.

An international panel selected the best papers from the first ten ICAD conferences, and reprinted them in a special issue of ACM Transactions on Applied Perception. This article was included in this publication.

Publication [P7]

T. Lokki, and M. Gröhn.
Navigation with auditory cues in a virtual environment.
IEEE Multimedia, Vol. 12, Nr. 2, pages 80–86, Apr.–Jun. 2005.

The findings of this article included the results of the two navigation experiments. The audio-visual navigation experiment is the same as in publication [P6], but the results were analyzed more in detail. Differences reported in [P6] were found to be statistically significant.

Totally new results are provided based on the comparison of carefully designed auditory cues and the basic auditory cue used in the first experiment. Navigation performance increased with auditory cues providing additional elevation and distance information.

These experiments are based on the author's code implemented for the audio-visual navigation experiment. The author ran the audio-visual navigation tests and made the basic statistical analysis described in the publication [P6]. Tapio Lokki ran the auditory cue comparison experiment and analyzed

both experiments more thoroughly. We both contributed to the discussion and conclusions.

Publication [P8]
M. Gröhn, T. Lokki, and T. Takala.
An orientation experiment using auditory artificial horizon.
In *Proceedings of the International Conference on Auditory Display (ICAD 2004)*, Sydney, 6.-9. Jul. 2004, 6 pages.

Results of the orientation experiments were included in this publication. We designed three different auditory artificial horizons to provide orientation information to users. We compared the subject's performance with these three and then without auditory cues. Subjects performed better with any auditory artificial horizon than without it. There was no significant difference in accuracy between the three auditory cues.

The author was responsible for this experiment and performed the majority of the work. Tapio Lokki and Tapio Takala assisted in test design.

Bibliography

[1] J. Blauert. *Spatial Hearing, The psychophysics of human sound localization.* The MIT Press, Cambridge, MA, 1997.

[2] R.S. Kalawsky. *The Science of Virtual Reality and Virtual Environments*. Addison-Wesley, Cambridge, UK, 1993.

[3] G. Burdea and P. Coiffet. *Virtual Reality Technology*. Wiley-Interscience, 2003.

[4] W.R. Sherman and A.B. Craig. *Understanding Virtual Reality: Interface, Applicationa, and Design*. Morgan Kaufmann, 2003.

[5] J. Pressing. Some perspectives on performed sound and music in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(4):482–504, 1997.

[6] S.A. Brewster, P.C. Wright, and A.D.N. Edwards. A detailed investigation into the effectiveness of earcons. In G. Kramer, editor, *Auditory Display: Sonification, audification and auditory interfaces.*, pages 471–498. Addison-Wesley, 1994.

[7] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, W. Tecumseh Fitch, and S. Tipei. *Sonification Report: Status of the Field and Research Agenda.* ICAD, 1999.

[8] J.K. Hahn and H. et al Fouad. Integrating sounds and motions in virtual environments. *Presence: Teleoperators and Virtual Environments*, 7(1):67–78, 1998.

[9] W.W. Gaver. *Everyday Listening and Auditory Icons*. PhD thesis, University of California at San Diego, 1988.

[10] W.W. Gaver. Using and creating auditory icons. In G. Kramer, editor, *Auditory Display: Sonification, audification and auditory interfaces.*, pages 417–446. Addison-Wesley, 1994.

[11] S.A. Brewster. *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*. PhD thesis, University of York, 1994.

[12] E.D. Mynatt. Auditory representation of graphical user interfaces. In G. Kramer, editor, *Auditory Display: Sonification, audification and auditory interfaces.*, pages 533–556. Addison-Wesley, 1994.

[13] P.A. Lucas. An evaluation of the communicative ability of auditory icons and earcons. In *Proceedings of the International Conference on Auditory Display'94*, pages 121–128, Santa Fe, NM, Nov 1994.

[14] F.L. Wightman and D.J. Kistler. Factors affecting the relative salience of sound localization cues. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 1–23. Lawrence Erlbaum Associates Inc., 1997.

[15] V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 62, 2001.

[16] R.O. Duda. Elevation dependence of the interaural transfer function. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 49–75. Lawrence Erlbaum Associates Inc., 1997.

[17] J.C. Middlebrooks. Spectral shape cues for sound localization. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 77–97. Lawrence Erlbaum Associates Inc., 1997.

[18] D.W. Grantham. Auditory motion perception: Snapshots revisited. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 295–313. Lawrence Erlbaum Associates Inc., 1997.

[19] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.

[20] K. Saberi and E.R. Hafter. Experiments on auditory motion discrimination. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 315–327. Lawrence Erlbaum Associates Inc., 1997.

[21] D.R. Perrot and T.Z. Strybel. Some observations regarding motion without direction. In R.H. Gilkey and T.R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 275–294. Lawrence Erlbaum Associates Inc., 1997.

[22] T. Takala and J. Hahn. Sound rendering. *Computer Graphics*, SIGGRAPH'92(26):211–220, 1992.

[23] D. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge, MA,, 1994.

[24] J. Huopaniemi. *Virtual acoustics and 3-D sound in multimedia signal processing*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 53, 1999.

[25] ITU-R Rec.BS.775-1. Multichannel stereophonic sound system with and without accompanying picture. Technical report, International Telecommunication Union, Geneva, Switzerland, 1992-1994.

[26] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustic Society of America*, 93(5):2764–2778, May 1993.

[27] D.G. Malham and A. Myatt. 3-D sound spatialization using ambisonics techniques. *Computer Music Journal*, 19(4):58–70, 1995.

[28] V. Pulkki and T. Lokki. Creating auditory displays with multiple loudspeakers using VBAP: A case study with DIVA project. In *Proceedings of the International Conference on Auditory Display'98*, Glasgow, Scotland, Nov 1998.

[29] F.L. Wightman and D.J. Kistler. Localization of virtual sound sources synthesized from model HRTFs. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'91)*, New Paltz, NY, 1991.

[30] E.M. Wenzel. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1):80–107, 1992.

[31] E. Wenzel, M. Arruda, D. Kistler, and S. Foster. Localization using nonindividualized head-related transfer functions. *Journal of Acoustic Society of America*, 94:111–123, 1993.

[32] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert. An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. *ACUSTICA acta acustica*, 86:1046–1053, 2000.

[33] D.R. Begault, E.M. Wenzel, and M.R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, 49(10):904–916, 2001.

[34] J. Sandvad. Dynamic aspects of auditory virtual environments. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4226.

[35] R.L. Martin, K.I. McAnally, and M.A. Senova. Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society*, 49(1/2):14–22, Jan./Feb. 2001.

[36] V. Pulkki. Localization of amplitude-panned virtual sources I: Stereophonic panning. *Journal of the Audio Engineering Society*, 49(9):739–752, Sept. 2001.

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

[38] V. Pulkki and T. Hirvonen. Localization of virtual sources in multi-channel audio reproduction. *IEEE Transactions on Speech and Audio Processing*, accepted for publication 2004.

[39] E. Wenzel. Effect of increasing system latency on localization of virtual sounds. In *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, pages 42–50, Rovaniemi, Finland, April 10-12 1999.

[40] Y. Suzuki, T. Yokoyama, and T. Sone. Influence of interfering noise on the sound localization of a pure tone. *Journal of Acoustic Society of Japan*, 14(5):327–339, 1993.

[41] P. Minnaar, S.K. Olesen, F. Christensen, and H. Møller. Localization with binaural recordings form artificial and human heads. *Journal of the Audio Engineering Society*, 49(5):323–336, May 2001.

[42] C. Jin, A. van Schaik, V. Best, and S. Carlile. Perceptual spatial-audio coding. In *Proceedings of the International Conference on Auditory Display 2003*, pages 255–258, Boston, MA, USA, July 6-9 2003.

[43] S.D. Steck and H.A. Mallot. The role of global and local landmarks in virtual environment navigation. *Presence: Teleoperators and Virtual Environments*, 9(1):69–84, 2000.

[44] R.A. Ruddle and S.J. et al Payne. Navigating large-scale 'desk-top' virtual buildings: Effects of orientation aids and familiarity. *Presence: Teleoperators and Virtual Environments*, 7(2):179–193, 1998.

[45] J.L. Chen and K.M. Stanney. A theoretical model for wayfinding in virtual environments: Proposed strategies for navigational aiding. *Presence: Teleoperators and Virtual Environments*, 8(6):671–686, 1999.

[46] R.P. Darken and J.L. Sibert. A toolset for navigation in virtual environments. In *UIST '93: Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 157–165, New York, NY, USA, 1993. ACM Press.

[47] J.M. Loomis and R.G. et al Golledge. Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments*, 7(2):193–204, 1998.

[48] S. Holland, D. Morse, and H. Gedenryd. AudioGPS: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6(4):253–259, 2002.

[49] P. Rutherford and D. Withington. The application of virtual acoustic techniques for the development of an auditory navigation beacon used in building emergency egress. In *Proceedings of the International Conference on Auditory Display 2001*, pages 144–149, Espoo, Finland, Jul 2001.

[50] T. Lokki, M. Gröhn, L. Savioja, and T. Takala. A case study of auditory navigation in virtual acoustic environments. In *Proceedings of the International Conference on Auditory Display 2000*, pages 145–150, Atlanta GA, Apr 2000.

[51] B. Walker and J. Lindsay. Effect of beacon sounds on navigation performance in a virtual reality environment. In *Proceedings of the International Conference on Auditory Display 2003*, pages 204–207, Boston, MA, Jul 2003.

[52] B. Walker and J. Lindsay. Auditory navigation performance is affected by waypoint capture radius. In *Proceedings of the International Conference on Auditory Display 2004*, Sydney, Australia, Jul 2004.

[53] D.A. Bowman and E.T. Davis. Maintaining spatial orientation during travel in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 8(6):618–632, 1999.

[54] J.R. Lackner and P. DiZio. Spatial orientation as a component of presence: Insights gained from nonterrestrial environments. *Presence: Teleoperators and Virtual Environments*, 7(2):108–116, 1998.

[55] C. Cruz-Neira, D. Sandin, and T. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. Anaheim, California, USA, 1993.

[56] J. Jalkanen. *Building a spatially immersive display - HUTCAVE. Licenciate Thesis*. Helsinki University of Technology, Espoo, Finland, 2000.

[57] J. Hiipakka, T. Ilmonen, T. Lokki, and L. Savioja. Sound signal processing for a virtual room. In *Proc. X European Signal Processing Conference (EUSIPCO 2000)*, Tampere, Finland, September 2000.

[58] J. Hiipakka, T. Ilmonen, T. Lokki, M. Gröhn, and L. Savioja. Implementation issues of 3D audio in a virtual room. In *Proc. SPIE*, volume 4297B, San Jose, California, January 2001.

[59] P. Zahorik, D.J. Kistler, and F.L. Wightman. Sound localization in varying virtual acoustic environments. In *Proceedings of the International Conference on Auditory Display'94*, pages 179–186, Santa Fe, NM, Nov 1994.

[60] A.J. Benson. Spatial disorientation – a perspective. In *RTO HFM Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures*, La Coruna, Spain, 15-17 April 2002. Published also in RTO-MP-086.

[61] MathWorks. Documentation for mathworks products, release 14 with service pack 1, matlab toolboxes, statistics toolbox. `http://www.mathworks.com/`, Nov. 2004.

[62] A. Bregman. *Auditory Scene Analysis: The Perceptul Organization of Sound*. The MIT Press, Cambridge, MA,, 1990.

[63] T.V.Tuyen and T. Letowski. Evaluation of acoustics beacon characteristics for navigation tasks. *Ergonomics*, 43(6):807–827, 2000.

[64] J.A. Ballas, D. Brock, J. Stroup, and H. Fouad. The effect of auditory rendering on perceived movement: Loudspeaker density and HRTF. In *Proc. Int. Conf. Auditory Display 2001*, pages 235–238, Espoo, Finland, 2001.

[65] M. Gröhn, T. Lokki, and L. Savioja. Using binaural hearing for localization in multimodal virtual environments. In *Proc. 17th Int. Congr. Acoust. (ICA 2001)*, volume IV, Rome, Italy, September 2001.

[66] B.N. Walker and G. Kramer. Mappings and metaphors in auditory displays: An experimental assessment. In *Proceedings of the International Conference on Auditory Display'96*, pages 71–74, Palo Alto, CA, Nov 1996.