Using 3D Audio Guidance to Locate Indoor Static Objects

Calle Håkansson and Samuel Sandberg

March 9, 2006
Abstract

We present a study investigating the use of 3D audio for navigation and location of objects in an indoor environment. In order to conduct the study, a custom-made 3D audio system was built based on a public-domain HRTF-library to playback 3D sound beacons through a pair of headphones. Our results indicate that 3D audio is indeed a qualified candidate for navigation systems, and may be especially suitable for environments or individuals where vision is obstructed, insufficient or unavailable. The experiment involved finding objects in an office environment. A comparison of 3D audio, map-based and unaided search shows significant differences between unaided and 3D audio search, as well as between unaided and map search. However, there were no significant differences between map and 3D audio search which is notable. The study suggests that elevation information should be implemented with special cues.

Keywords: 3D audio, headphones, headtracker, HRTF, navigation, wearable, auditive display, audio perception, RTLS, user study
Acknowledgements

We would like to thank our supervisors Fang Chen, Phillipas Tsigas and Niklas Elmqvist. We would also like to thank Ekahau Inc for the software license and The Regents of the University of California for the HRTF library.
Contents

1 Introduction .................................................. 8
  1.1 Purpose and hypothesis ................................. 9
    1.1.1 Limitations ........................................ 9

2 Background .................................................. 11
  2.1 Human hearing ........................................... 11
  2.2 Human perception ....................................... 11
  2.3 3D audio technologies .................................. 13
  2.4 Head Related Transfer Function ....................... 13
  2.5 Intelligibility 3D audio sources ..................... 14
  2.6 Other techniques for sonification of objects ......... 14
  2.7 Human audio cognition ................................ 15
    2.7.1 Cognitive workload and Dual task ................ 15
    2.7.2 Substrates for attention in neuroscience ....... 16
  2.8 Usable sounds for 3D-positioning .................... 17
  2.9 Physical localization of objects ..................... 17
    2.9.1 Positioning ....................................... 17
    2.9.2 Radio Frequency Identification .................. 18
    2.9.3 Ultra Wideband .................. 18
    2.9.4 Wireless Local Area Network ..................... 19

3 System development .................................... 20
  3.1 Ekahau Platform ........................................ 21
    3.1.1 Ekahau Engine 3.1 ................................ 21
    3.1.2 Ekahau Manager 3.1 ............................... 21
    3.1.3 Ekahau Client .................................... 23
  3.2 Hardware development .................................. 23
    3.2.1 Headtracker ....................................... 23
    3.2.2 Digital Compass - HMR 3300, Honeywell ........ 25
    3.2.3 Accelerometer - R2999M, Memsic ................. 26
    3.2.4 Microprocessor - BasicX-24, NetMedia .......... 27
  3.3 Audio software, AuPos ................................ 28
    3.3.1 Development progress .................................. 28
    3.3.2 Troubles during development ...................... 29
    3.3.3 Design ............................................. 29
    3.3.4 Performance ....................................... 30
    3.3.5 File management ................................... 30
    3.3.6 Headtracker interface ................................ 31
    3.3.7 Positioning system interface .................... 31
    3.3.8 Graphical User Interface .......................... 31
    3.3.9 Quality control .................................... 31

4 Pilot test ................................................. 33
  4.1 Specification of audio system ....................... 34
  4.2 Audio choice ............................................ 35
    4.2.1 Sound clip choice .................................. 35
    4.2.2 Audio playback function .......................... 35
5 Method
5.1 Participants ........................................ 38
  5.1.1 Test groups ................................... 39
5.2 Apparatus .......................................... 39
5.3 Materials ........................................... 41
5.4 Design ............................................. 41
  5.4.1 Minimizing biases ............................... 42
5.5 Procedure .......................................... 43
  5.5.1 Questionnaire ................................... 46

6 Results .................................................. 48
6.1 Experiment results ................................ 48
  6.1.1 Experimental assistant protocols ............ 48
6.2 Questionnaire results ............................... 49
6.3 Comments from participants ........................ 51
  6.3.1 3D audio condition .......................... 51
  6.3.2 Map condition ................................ 52
  6.3.3 No aid condition ............................... 52
6.4 Experiment assistant protocols ................... 52

7 Discussion ............................................ 53
7.1 Evaluating method ................................ 53
7.2 Evaluating experiment .............................. 53
  7.2.1 Performance .................................. 53
  7.2.2 Comments from participants ................. 54
7.3 Evaluating questionnaire ........................... 54
7.4 Evaluating 3D audio implementation ............. 54
7.5 Future work ......................................... 54
  7.5.1 Audio cues .................................... 54
  7.5.2 Headtracker ................................... 55
7.6 Suggested applications ............................ 55
  7.6.1 Working with 3D audio guidance ............ 55
  7.6.2 The development of an adaptive 3D audio system .... 56

A Appendix ............................................. 60
A.1 User scenarios ...................................... 60
A.2 Fraction papers ..................................... 61
A.3 Questionnaire ....................................... 62
A.4 Consent form ....................................... 63
A.5 Participant object order ............................ 64
A.6 HRTF selection instructions ....................... 65
List of Tables

1 Definitions .................................................. 10
2 ANOVA .................................................. 48
3 Post-hoc analysis ........................................ 49

List of Figures

1 Human ear .................................................. 11
2 Cone of confusion ........................................ 12
3 Human brain ............................................. 16
4 3D audio system overview ............................. 20
5 Ekahau communication ................................ 21
6 Ekahau manager 3.1 screenshot ..................... 22
7 Headtracker - inside ..................................... 23
8 Headtracker - outside ................................... 24
9 Azimuth and elevation ................................... 24
10 Schematics of headtracker ........................... 25
11 Memsic ................................................... 26
12 Angle measurement ..................................... 27
13 Layout of a .pos-file ................................... 30
14 GUI for choosing HRTF ............................... 31
15 Pilot participants ....................................... 33
16 HRTF angle measurements ............................ 34
17 Comparing bell spectrum and white noise spectrum .... 36
18 Volume cue .............................................. 37
19 Age representation ..................................... 38
20 Gender vs condition .................................... 38
21 Participants ............................................. 39
22 Card placement ......................................... 42
23 Cards .................................................... 43
24 Mounted helmet and headtracker ................... 44
25 Map of the IT-University ............................ 45
26 Box plot of conditions ................................ 49
27 Questionnaire results .................................. 50
28 Questionnaire results .................................. 50
29 User scenarios .......................................... 60
30 Fraction papers ......................................... 61
31 Questionnaire .......................................... 62
32 Consent form ........................................... 63
33 Participant object order ............................... 64
34 HRTF selection instructions ......................... 65
1 Introduction

An auditive display is a presentation media that uses audio as the information carrier. An auditive display displays the information to the user via the user’s hearing. You might say that a computer screen is a visual display and a radio is an auditive display. These two can of course be combined into an audio/visual display, for example a TV. The two common types of displays (audio and visual) have each benefits and drawbacks; A blind user cannot see a computer screen and a deaf user cannot listen to the radio. These two extremes are interesting to people without these disabilities because there exist a lot of situations where a user can be obstructed from using either ears or eyes. A fireman might not be able to read when inside a smoky building and a construction worker might not be able to hear the communication radio when working.

Auditive displays are nowadays used in several different areas and for several different reasons. In recent years, the 3D property of audio through headphones has been researched and methods of displaying audio that is percepted as coming from the real world outside the headphones have been developed. In the rest of this study, whenever 3D audio is concerned it is presupposed via headphones. 3D audio adds yet another dimension to the auditive display giving the end user a sense of spaciousness to more easily distinguish and locate sounds. This technology has been used in for example the music industry and in computer games.

Three dimensional audio signals can be used in many favorable ways, such as for shifting the attention of the user (warning signals), for aiding speech recognition of multiple simultaneous sources (the “cocktail party” effect), for rendering a virtual “audio map”, etc. It has been shown that the 3D audio property increase the overall performance of auditive displays.

The technology is helpful to, for example, a pilot who can hear where his wingman is or a soldier in the battlefield who can hear where his friends are, and thus not engage in friendly fire. For some occupations, such as fire-fighters and soldiers, real-time cooperation between the team members is an important factor for success. In order to achieve effective cooperation, having a clear awareness of the geographic locations of every member is critical. Maps (both static and dynamic) are one of the classic tools to help people gain such geographic overviews, yet they require visual input. For some users, this may not be possible: either the visual channel is unsuitable, i.e. looking at the map would distract the user and cause him to miss critical visual information in the surroundings, or the environment does not permit visual navigation aids (due to smoke, darkness, mud, etc). 3D audio also increase the verbal understanding of multiple simultaneous messages and can be used by operators to more easily distinguish a particular voice.

The 3D audio effect is very subtle and thus requires high quality audio with a carefully chosen audio feed. The processing of 3D audio is CPU intensive and have earlier limited the applications. In recent years 3D audio using headphones have become a more interesting field of research since hardware prices continue to drop.

There are several ways to achieve the 3D effect via headphones; the most delicate way is to process the audio through individualized HRTF filters that mimic the user’s ears, pinnae, torso and head [King and Oldfield, 1997] [Chen and Carlander, 2003]. The HRTF-technology thus creates a computer model of the hearing of a specific person; every human percept audio in different ways. Creating a HRTF is expensive and therefore it is somewhat troublesome to incorporate it in applications. However, an acceptable HRTF can most often be found in a small set of HRTFs.
1.1 Purpose and hypothesis

The downsides with 3D audio displays are that they require a headtracker device and hardware for playing the sounds. A headtracker is a device that sense the head movement of the user. Using it will increase the overall 3D perception and guarantee that the audio always is presented relative to the user. Another downside is that badly designed 3D audio presentation might interfere with ongoing mental processes, e.g. disturb the user. The hardware is really only a matter of money, but using too much mental effort for audio perception may lead to very poor over all task performance.

Since very few conducted experiments have focused on real life tests, the goal of this study was to provide data that has not yet been gathered concerning 3D audio. The purpose of this study was to investigate whether 3D audio is a useful alternative to visual navigation for locating static objects indoors. Previous research [Zhou et al., 2004] has shown good results when navigating inside a room. This study will however test if previous studies can be applied to a whole building.

There are some words in this study that might not be well-known or well-defined and are therefore gathered in table 1.

This report will first give an introduction to the theories of human audio perception and cognition. The theories are used when developing a 3D audio system. In the final sections of the report, the 3D audio system is used as a tool to answer the research question and the hypothesis. Finally, the results are presented and evaluated.

1.1 Purpose and hypothesis

The main goal for this study is to show if 3D audio is a useful tool for people in a real-life situation. This study can be used for guidance when building auditive displays and involves both an experiment to show differences between groups with different aids and a questionnaire to evaluate the participant experience of the aid.

The research question in this study is “Will 3D audio provide useful help to locate static objects indoor?”. It addresses whether 3D audio is a candidate for displaying information in real-life situations. This study will investigate the differences between auditive and visual displays to find out how they might differ.

The hypothesis is “It is possible to successfully navigate and find objects using 3D-audio information without the help of external visual aids”. The results will show performance comparisons which will be most helpful in future research of practical auditive displays.

1.1.1 Limitations

The study will not examine different ways to display auditive information, e.g. it will only display auditive information with 3D audio cues and volume cues via headphones. The experiment will only be conducted in one particular building which is assumed to represent a default building, which may introduce bias. Further more, the results are limited by the shortcomings of the custom built 3D audio system.
# INTRODUCTION

<table>
<thead>
<tr>
<th>Word</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headtracker</td>
<td>A hardware device for tracking the direction of the end user’s head.</td>
</tr>
<tr>
<td>Auditive display</td>
<td>A display that use audio as information media.</td>
</tr>
<tr>
<td>3D audio</td>
<td>Audio that is experienced as if the audio was played outside the headphones.</td>
</tr>
<tr>
<td>HRTF</td>
<td>A Head Related Transfer Function is an individual computer model of hearing.</td>
</tr>
<tr>
<td>AuPos</td>
<td>The software designed to play 3D audio. It was created solely for this experiment.</td>
</tr>
<tr>
<td>Ekahau</td>
<td>Company that provides software which can position network cards using WLAN.</td>
</tr>
<tr>
<td>Crossfade</td>
<td>Fade out one sound and fade in another sound simultaneous to get a smooth audio transition between the two sounds.</td>
</tr>
<tr>
<td>Headphones</td>
<td>A pair of speakers that is placed in close proximity to the ears.</td>
</tr>
<tr>
<td>Earphones</td>
<td>Earphones are small headphones put in the auditory canal.</td>
</tr>
<tr>
<td>Broadband spectrum</td>
<td>A sound that has a wide range of frequencies.</td>
</tr>
<tr>
<td>sound</td>
<td></td>
</tr>
<tr>
<td>White noise</td>
<td>White noise has equal sound pressure level (volume) for all frequencies.</td>
</tr>
</tbody>
</table>

Table 1: Definitions.
2 Background

This section gives a short presentation of the spatial audio theories used in this study.

2.1 Human hearing

The most evident part of the human hearing is the outer ear and the pinnae (often called “the ear”) which collects the sounds and transports them through the auditory canal. The sound is converted to vibrations by the ear drum and is amplified in the middle ear by the malleus, incus and stapes. In the inner ear, the sound vibrations are converted to fluid pressures in the cochlea and receipted by hair cells that send nerve impulses to the brain.

Figure 1: Human ear.

2.2 Human perception

3D audio theory is based on the human perception and it is essential to know some basic audio and physiological properties to understand the problems and challenges of 3D audio displays. The human location accuracy is about 10 degrees in the horizontal plane [Yuzhu et al., 1998] and worse for the vertical plane.

A human has two ears but the hearing is not only based on the sound that comes in through the pinnae; sound also reach the eardrum through the body, e.g. the torso and the head [Chen and Carlander, 2003]. That is important for the 3D perception of audio because the sound travels in different paths through the body depending on the
direction it was originating from. Basic 3D audio perception is however more easy to grasp. The two major features are inter-aural time differences (itd) and inter-aural intensity differences (iid). The itd is based on that a sound will reach the ears with a time difference and that the time difference is used to calculate the direction of the sound. The difference between the ears on a normal sized head is approximately 0.8ms. Itd also encounters another problem when a high frequency sound is played; if the phase difference is a whole wave-length it can be interpreted as no phase difference and thus not be correct displayed (the direction to the sound source is misinterpreted). Therefore, itd is mainly used for frequencies below 1500Hz.

Inter-aural intensity difference on the other hand uses the difference in intensity, e.g. volume, to perceive the sound in three dimensions. The sound will sound louder in the ear nearest the sound source, not so much for the distance between the ears (only important if the sound is very near, e.g. a whisper) but for the shielding effect of the head. This is true for high frequencies (above 3000Hz) and is called the head-shadow effect. For low frequencies the sound pressure is approximately the same around the head and the iid is not applicable.

There is unfortunately an effect called the cone of confusion (see figure 2) which means that the sound can originate from the border of the cone base and still fulfill the requirements of itd and iid, e.g. the distance and time delay to the ears are same. The cone of confusion is evident, though it is not completely true in practice since all humans are asymmetric and small deviations from the theoretical approach will generate audio cues that will be interpreted. Due to both iid and itd it is important to use a broadband spectrum sound so both can be used for interpreting. However, as stated before, the sound does not only come through the ears. It shows that the high frequencies are most important for spatial audio localization due to reflections in the pinnae [Kistler and Wightman, 1992]. Frequencies in midrange 1500-3000Hz are hard to localize [Chen and Carlander, 2003] since they cannot efficiently take advantage of the above techniques to discriminate the location. An example of sound with these frequencies is human speech.

Figure 2: Cone of confusion. The cone base border shows from where a sound might be interpreted to originate.
In addition to itd and ild, humans use dynamic cues to localize objects. Often sounds are not static, and more often the listener is not static. By moving and rotating the head, the dynamic cues remove the cone of confusion and thus render an appropriate spatial presentation.

Since childhood, the hearing is trained to learn spectral patterns in familiar sounds. Humans learn to hear how a sound is sounding at different distances. This experience can help when localizing sounds, especially for hearing distance [Chen and Carlander, 2003]. As humans listen after dynamic cues, the need for user interaction is evident. A head-tracking device improves headphone-localization accuracy to equal free-field [Langendijk and Bronkhorst, 1999] by letting the user move and rotate the head while displaying the sounds at the same virtual place.

### 2.3 3D audio technologies

When using a pair of headphones, the sounds must be altered in a way that mimic the physical way a sound is perceived since headphones only have one sound source per ear. A few of the common technologies/APIs are presented in short:

**HRTF** is the most delicate 3D audio technology. It is an individual physical representation of the hearing. The best result is yielded when the user have a custom made HRTF. It is expensive to create individual HRTFs so often a best match is found in a HRTF-set.

**Lake Technology Limited** made some breakthroughs in the signal processing and psychoacoustics of 3D audio in the early 1990s and made a commercial system without the need for custom made HRTFs.

**Dolby Headphone** is a system that can play 5.1 multichannel surround sound in a pair of headphones with virtual speakers. It is based on the Lake technology.

**DirectSound3D** is a 3D audio API from Microsoft which includes support for doppler-effects, distance attenuation et.c.

**EAX** from Creative Labs has the most advanced physical sound simulation for computer games with features as object obstruction, environment morphing et.c.

### 2.4 Head Related Transfer Function

The primary technique for rendering virtual 3D audio via headphones is the head related transfer function (HRTF). The HRTF is a set of impulse responses of the ears from certain directions which contains a detailed measurement of the acoustic features of the ears, pinnae, torso and head. The HRTF makes it possible to play a sound that sounds as if it originated from a direction recorded as an impulse response. The HRTF is therefore individual, which is a problem when implementing the technique in a general application. The measurement of a HRTF is time consuming and requires expensive equipment. A small microphone is placed in the auditory canal to measure the sound intensity inside the ear. Since every impulse response only is valid from a certain direction, it is necessary to carry out multiple measures. It is recommended [Langendijk and Bronkhorst, 1999] to have a six degree spatial difference between the
impulse responses to not introduce audible interpolation cues, e.g. no audible transitions between two adjacent impulse responses can be heard.

There exist libraries with HRTFs that can be used for research in 3D audio. The major problem seems to be the effort it takes to determine an appropriate HRTF for a test subject. However, Seeber and Fastl (2003) have developed a method for subjective selection of HRTF that takes ten minutes per subject. The method consists of two steps; in the first step the subjects choose five HRTFs out of twelve following certain criteria’s and rank them. In the second step the test subjects rank the remaining five HRTF following more specific criterias. The result of this study shows that eight out of ten subjects choose the optimal or nearly optimal HRTF. The selection instructions can be read in appendix A.6 and they are translated from a mail communication with Bernhard U. Seeber (personal communication, November 4, 2005).

### 2.5 Intelligibility 3D audio sources

Navigation with both auditory information and visual information are easier and faster when a user is guided through a path, than if only one source of information is used [Lokki and Gröhn, 2005]. The audio information is most useful for approximating the direction though it is completely possible to navigate entirely by audio cues. Experienced fighter pilots prefer 3D-sound in wingman communication and warning signals compared to common non-localized audio. They also prefer sounds to direct them to a physical device [Lee, 1993].

Humans are able to understand two verbal messages simultaneous with a certain degree of spatial separation and if the sentences are short. Localization is, according to Chen (2003), slightly more accurate if the verbal messages are presented in the front or back side of the subject than on the left or right sides. In a dual task study performed in a car simulator, subjects recognized audible information more easily if the information was presented at spatially different places and with different timbre for each location in spatial space. This also lowered the mental workload for the subjects [Takao et al., 2002].

Research suggest that humans ability to detect a verbal target and localize it decrease by the number of talkers present. In a study, the correct detection was less than 75% when three or more talkers were present [Brock et al., 2003]. In a study of the neural base for sound source localization Palomäki et al. [Palomäki et al., 2005] found that although both hemispheres in the brain processes audio localization, the auditory cortices in the right hemisphere respond to, and are extra sensitive to auditive spatial localization. Contralaterally presented stimuli are processed more and faster than ipsilateral stimuli.

### 2.6 Other techniques for sonification of objects

There are alternative ways to HRTF for sonification of objects. Data can be incorporated into the sounds in different ways. One successful method for easy detection of corners and depth continuities as well as the size of the surrounding space, is mapping pitch and loudness to distance information. Milios et al (2003) tested two systems for blind people based on pitch-mapping. The proportional mode, where the range was directly converted into a pitch, was intuitive and worked best with complex surroundings and the derivative mode, where the change in range was converted into pitch, worked best with objects close in front of a flat surface. Users expressed an opinion
that a combination of both modes would be more informative than either would be alone. Modulating pink noise with gain + pitch to include elevation information resulted shorter path length, e.g. the subjects could locate the objects faster than if no pitch modulation was applied to the noise. Auditory navigation is almost as easy as visual navigation when sound signals are carefully designed [Lokki and Gröhn, 2005].

2.7 Human audio cognition

In this section it is explained why sound is a usable option for presentation of information if the semantics of the sound is carefully considered so it does not interfere with upcoming events in the environment it is intended to be used in.

2.7.1 Cognitive workload and Dual task

Cognitive workload is mainly associated with the demands of a task; how difficult one think it is to solve certain kinds of problems. Now and then people notice that certain peripheral stimulus, sounds for example, affects a certain task that they are focused on more than others. These kinds of peripheral stimuli is said to share more mental resources with the task that they are engaged in than other stimulus [Banbury et al., 2001], e.g. they are represented in a similar way in the memory. To explain the phenomena further, it is not the actual carrier of the sensation (vision, hearing et cetera) that is the problem but the message delivered. If it is a verbal message it does not matter if it is delivered via ones ears or eyes, it is equal disturbing if the task at hand is in the verbal domain. Sometimes you have to attend to different kinds of information at once (dual task). An example of this is to focus on both instruments and radio communication in a cockpit environment. In these circumstances it is important to try to minimize the workload as much as possible. A too demanding environment can cause disasters like the plane tragedy in Tenerife [Weick, 1990]. Research about workload has occupied many scientists including Yeh and Wickens (1988). Yeh and Wickens studied performance and subjective workload using a theory, that among other things, states that performance is related to level of similarity between two tasks and the difficulty of the tasks. One of their findings was that “performance is worse when two tasks compete for common resources” ... “than when two tasks demand different resources” [Yeh and Wickens, 1988]. This was also found in an experiment conducted by Wickens and Liu (1988) where they compared verbal and spatial decision tasks. Their results provide support for the importance of the dichotomy between verbal and spatial processing codes in accounting for task interference. Similar results were also found in a study by Gillie and Broadbent (1989). They used a computer game and disturbed the participants with interruptions of different duration, different level of likeness with the ongoing task in the game and different level of difficulty. Like Yeh and Wickens (1988) their findings support the notion that processing similar material significantly decrease performance on the primary task. A conclusion that can be drawn from the facts presented above is that it is possible to attend to visual and auditory information at the same time as long as the information delivered is not semantically alike, for example it would be hard to write a report on the computer while you have to take part in a phone conversation. One further example is to listen to a voice navigation system in a car where the system tells the driver to turn left at the same time someone in the backseat mention the word right in a conversation. This may cause the driver to mix the sources for information and turn right instead of left [Wickens et al., 2004].
2.7.2 Substrates for attention in neuroscience

The feeling of being in control over our thoughts and actions is called executive attention [Nyberg, 2002]. One other type of executive attention is attentional filtering of which the cocktail party phenomenon is a famous example. The cocktail party phenomenon means that one selectively can tune in a specific sound source while filtering out other stimuli like the person talking in front of you [Coren et al., 1999]. Other types of attention are for an example to attend to a warning signal or to see something moving in one's corner of the visual hemifield and hereby switch attention. This kind of attention is involuntary. We respond to the signals no matter if we want to or not. Sounds for instance seem to have automatic access to attention. Research has shown that if a person's name is heard that person automatically redirects his/her attention to the sound source (Neville Moray, 1959, cited in [Gazzaniga et al., 1998]). In executive attention, a part of the frontal lobe called cingulate gyrus is especially active.

Dual attention is very demanding on the executive attention. When two tasks are executed at the same time, the performance on both tasks are usually reduced. This is especially true when the two tasks are similar. Semantically alike stimuli are processed in the same area of the brain (see figure 3) in higher level of processing. Linguistic information for instance is recoded in Wernickes area in the temporal lobe [Nyberg, 2002], no matter if the linguistic information is seen, heard or touched. The activation in cingulate gyrus is increased during dual tasks, also other parts of the frontal lobe shows increased activation. In fact a bigger part of the conscious parts of the brain (neocortex) is activated during dual attention than the total sum of the areas activated when performing each activity separately [Nyberg, 2002]. The level of activation is also depending on how much one has been trained to carry out a task. Automated tasks, for instance driving a car for a cabdriver, require much less attention than tasks one is not used to carry out.

Figure 3: The lobes of the human brain. Clockwise from the upper left corner: Frontal lobe, Parietal lobe, Occipal lobe and Temporal lobe. Image courtesy of [University of Utah health sciences center, 2006]

![Brain lobes image](image_url)
2.8 Usable sounds for 3D-positioning

It is important to choose good audio signals for localization purposes. Broadband, impulsive, longer duration signals are easy to locate while low frequent, low envelope signals are difficult to locate [Yuzhu et al., 1998]. Test labs often use broadband noise (in bursts) but there are other aspects that also are important localization features in the real world, e.g. familiarity with the signal and if the perceived sound is pleasing.

Based on the theories presented in section 2.2 (page 11) it is important to have a broadband audio signal to perceive correct direction. It is also recommendable to use familiar sounds and preferably in bursts. If a user shall listen to the same sound for a long time, it is important to use a sound which will not negatively affect the user. In most lab test white noise bursts are used because they are believed to give the user the highest 3D experience since the white noise contain all frequencies. This will probably not be suitable in a real environment during long sessions due to fatigue effects.

2.9 Physical localization of objects

One other aspect when using 3D-audio is to map the physical relation between the user and the virtual or physical object. Some aspects to consider are if the object is static or movable, in what direction the user is turned etc. Zhou et al. (2004) conducted a combined 3D audio and Augmented Reality (AR) study, where the participants used a camera attached to a head mounted display (HMD) to identify objects. To know the spatial relation between the user and the objects Zhou and colleagues used a combination of gyroscopes, accelerometers and ultrasonic sensors. This yielded a maximum resolution of 15 mm. Some studies described above uses advanced technique to explore the area of 3D-audio e.g. [Takao et al., 2002, Zhou et al., 2004]. Gaye (2002) approached 3D-audio in a different way, Gaye tried out a 3D-audio system that could be adapted to different levels of CPU power and size of computer memory to enable usage in different devices e.g. handheld computers, PCs etc. The study compared different models for moving around a virtual sound source. Gaye used headphones, a headtracker and a PC in this study. Goodrum et al. (2005) conducted a study to analyze a tool tracking device at a work site using active RFID technology. The technology used did not deliver the performance needed for such a system; only tools in a 25m radius from the user could be monitored.

“...significant value can be gained, however, if future research addresses how affordable technology could provide direction and range data for tagged assets located throughout a construction job site.”
[Goodrum et al., 2005]

2.9.1 Positioning

To be able to test 3D audio in an applied setting indoors an indoor positioning system had to be used. A few technologies were found interesting: Radio Frequency Identification (RFID), Ultra Wideband (UWB) and finally Wireless Local Area Network (WLAN). First signaling with narrowband transmitters (ER400TRS Transceiver, Easy radio) were tested but was however rapidly rejected because of the effort it would take to build a fully working tracking system from that technology. The purpose of this study is to investigate 3D sound and see if it can be applied to indoor settings for localization of objects, not to build a tracking system. The tracking system is only an aid to test 3D audio. Global Positioning System (GPS) that is one of the most commonly
used technologies was not interesting in this project since it is dependent on a clear view to the sky for accurate positioning. Positioning through WLAN was eventually found to be the best candidate for this study. The rationale for that decision was:

- It was cheaper since the technology was already up and running in the test area (IT-University, Chalmers University of Technology, Gothenburg)
- It was found to be more interesting to test 3D audio on technical infrastructure that is available in many buildings (schools, airports, museums, train stations etcetera) and see if those systems were sufficient for 3D audio positioning.
- WLAN was found suitable to use for positioning and had adequate spatial precision for this study using the Ekahau system.

The 3D audio system developed in this study is easy to adapt to other positioning technologies. Below a brief overview of the techniques Radio Frequency Identification (RFID), Ultra Wideband (UWB) is presented. Wireless Local Area Network (WLAN) and how it was used is described in more detail since that was the technique chosen for this study.

### 2.9.2 Radio Frequency Identification

Radio Frequency Identification (RFID) can be divided into two subgroups, passive and active RFID. Passive RFID means that the RFID tag receives power externally (from a reader) to send data to the reader. Following this the tag does not need batteries. Passive tags have small data storing capacity and short reading range; up to 4 meters with a stationary reader and less than 16 centimetres using a handheld reader, for instance personal digital assistant (PDA) [Goodrum et al., 2005]. The major pros of passive RFID are that the tags are small and cheap. Active RFID on the other hand contains batteries and can transmit data without having to induce external energy. Active RFID have further a longer reading range than passive RFID, up to 25 meters and more data storing capacity [Goodrum et al., 2005]. Goodrum et al. (2005) conducted a study where the purpose was to build a tool tracking and inventory system using Active RFID and a PDA. The results from the study infer among other things that RFID is not suitable for precise location of objects,

“Currently, active RFID tags do not provide directional or range data from the reader to the tag. As experienced during the field trials, the reader would identify the tagged tools that were within range of the reader but without providing information on the bearing or distance to the tagged tool” [Goodrum et al., 2005]

Active RFID would probably not be suitable for this study where spatial location and direction to target are important factors.

### 2.9.3 Ultra Wideband

Previous studies [Fontana and Gunderson, 2002] [Hirt, 2003] [Gu and Taylor, 2003] [Oppermann, 2004] has shown that UWB technology is capable of location accuracies down to a few centimetres indoors. One of the major advantages using UWB indoors is that UWB technology is comparatively insensitive to Multipath problems which are common indoors. Multipath means
2.9 Physical localization of objects

“Involving or relating to signals which arrive at a receiver having come from the same source by different routes”
[Oxford English Dictionary, 2006]

One other forthcoming with UWB is that it has good penetration capability through different materials e.g. walls and other obstructions [Fontana et al., 2003]. One localization system using UWB is the PAL650 (Multispectral solutions incorporated). The PAL650 works as follows; three or more receivers are placed at known coordinates around the targeted area. One reference tag is placed inside the target area at a known location. Tags can be traced inside the target area through a “central processing hub” [Fontana, 2004].

2.9.4 Wireless Local Area Network

Wireless Local Area Network (WLAN) was found to be a suitable technology for tracking objects indoors in this study. A short description of the technology follows below. After that a description of the WLAN software used is described.

“A wireless LAN or WLAN is a wireless local area network that uses radio waves as its carrier: the last link with the users is wireless, to give a network connection to all users in the surrounding area. Areas may range from a single room to an entire campus. The backbone network usually uses cables, with one or more wireless access points connecting the wireless users to the wired network” [Wikipedia The free encyclopedia, 2006]
3 System development

The 3D audio system used in this study is described in detail below. A system overview can be seen in figure 4.

Figure 4: An overview of the 3D audio system used in this study.
3.1 Ekahau Platform

To be able to spatially locate one or several users moving around in a building this 3D audio project uses Wireless Local Area Network (WLAN), which is a part of the available technical infrastructure at the test location. The location system used is commercial software from Ekahau (Ekahau Inc.). The Ekahau system is based on probability theory, e.g. if a Wi-Fi adapter on a laptop records certain signal strengths from Wi-Fi access points XYZ the adapter is probably at location in the building see figure 5 for an overview of the Ekahau products used in this project.

Figure 5: Ekahau communication.

3.1.1 Ekahau Engine 3.1

The Ekahau Positioning Engine is according to Ekahau “the only software based real-time location system in the market” [Ekahau Inc, 2006a]. The system is calibrated through measuring the signal strength at several locations in a Wi-Fi (802.11a/b/g) network. The system needs to be in contact with at least three Wi-Fi (802.11a/b/g) access points at each measuring point. To calibrate the Ekahau Positioning Engine you need to have a supported Wi-Fi (802.11a/b/g) adapter installed on your laptop [Ekahau Inc, 2006b].

3.1.2 Ekahau Manager 3.1

Ekahau Manager can be looked upon as the GUI for Ekahau Engine, see figure 6. In Ekahau Manager the user can make several settings to the Ekahau Engine. Below the settings used in this project are listed [Ekahau Inc, 2006c]:

- Upload maps over different floors in a building
- Add Tracking Rails to a map. Calibration points can only be measured standing on a Tracking Rail. By adding tracking rails different floors in a building can be
connected to each other so that an object can be traced when it moves between different floors.

- Performing Site calibration. To be able to track an object you have to measure Wi-Fi signal data from various map locations. This is done by rotating the laptop at different directions standing at a specific place in a building. It is important that the representation on the map corresponds to the real world. The ideal site calibration is done by measuring Wi-Fi signal data every three meters in a grid. In this project totally 348 calibration points were measured on three floors. The spatial accuracy is estimated to be $0.1 - 4.5$ meters. The spatial accuracy varies between floors and between different sites on each floor.

- Logical areas. Three logical areas are used in this project, floor two, floor three and floor four. A logical area is used to determine if an object is within an area or not. This setting is especially helpful if you track several objects at the same time at different floors in a building.

- Selecting access points. The Ekahau Manager lists all Wi-Fi access point available. Some of these access points may be from another network nearby. To avoid using all access points available when calibrating the system you are able to choose which ones you want to use. Doing that reduces the chance that chosen access points are moved without your awareness which may cause disturbance in the calibrated system.

- Set the map scale so it matches the real world. By using the measure tool one can choose a distance on the map and then measure the real world distance. By
3.2 Hardware development

• Save positioning models to the engine e.g. the maps, rails, measuring points etcetera are saved to the positioning engine.

• Loading positioning models. When the manager program starts up you may choose which positioning model you want to use.

3.1.3 Ekahau Client

The program Ekahau Client has to be installed on every wireless unit that one wish to track. In Ekahau Client it is possible to choose which engine you grant for positioning of that Client device. By doing that the granted engine is allowed to retrieve signals for positioning of that client device [Ekahau Inc, 2006c].

3.2 Hardware development

3.2.1 Headtracker

A custom-made headtracker (see figure 7 and figure 8) was built in order to get direction data and tilt angle data from a user’s head, e.g. to know where the user’s nose was pointing (see figure 9).

Figure 7: The inside of the headtracker.

The main parts of the headtracker are a microprocessor (BasicX-24, NetMedia), an electronic compass (HMR 3300, Honeywell) and an accelerometer (R2999M, Memsic). The head tracker has an update frequency of about 8Hz, and even though an update frequency of at least 14 Hz is recommended (brungart-04) and some 3D audio systems use up to 100Hz, the overall performance of the system was found acceptable. A wiring diagram of the final version of the headtracker can be seen in figure 10.

The data from the headtracker (azimuth from compass and elevation from accelerometer, see figure 9), is sent to a laptop that is running the positioning- and 3D audio application. This is managed through the communication (COM) port. Besides
the apparatus mentioned above ultrasound and gyros were considered to be parts of the headtracker. Ultrasound was considered as an alternative to get range data from the environment, to work as backup to the rest of the system. The rationale for these thoughts was that if the compass and accelerometer indicates a movement, but if the data from the ultrasound does not indicate movement, maybe no movement has taken place. These thoughts were however soon rejected since people would be walking around the user thus disturbing the ultrasound sensor, and because of the lack of range precision of the ultrasound gear tested (SRF04 - Ultra-Sonic Ranger, Devantech). The reason for not choosing gyroscopes is the drift. Affordable gyros drift up to several degrees per minute [Foxlin, 1996]. This means that a user of such a system might experience objects floating, for an example the objects seems to float away in an audio system where sounds shall represent certain directions to objects. One aspect considered when the final version of the headtracker was already up and running was to
3.2 Hardware development

Figure 10: Schematics of headtracker.

combine the precision of the compass with the speed of data signals and robustness to outer interferences of a gyroscope. One solution could be to use the gyro for perhaps 30 seconds and then calibrate it using the compass. Doing so would compensate for the compass sensitiveness to magnetic fields and the gyroscopes tendency to drift over time. Below a description of the apparatus used to build the headtracker.

3.2.2 Digital Compass - HMR 3300, Honeywell

Below some information gathered from [Honeywell, 2006]:

- Heading Accuracy, 0.1 ° Resolution
- ±60 ° Tilt Range, Pitch and Roll
- Compensation for Hard Iron Distortions, Ferrous Objects, Stray Fields
• $-40 \degree C$ to $85 \degree C$ Operating Temperature Range

The compass is used for azimuth data, see figure 9. The compass works even if the it is tilted which is crucial in this study e.g. the compass must work even if the participants tilt their heads in different directions. The compass has an update frequency of 8 Hz. This is a limitation in this study and 8 Hz is the maximum speed the headtracker can send data to the 3D audio- and positioning system. Pilot tests (see section 4 on page 33) however showed that the headtracker was fast enough to be of practical use. The compass has built-in accelerometers to compensate for tilt. The data from the accelerometers were however not used as frontal elevation data in this study since it was found to further slow down the data communication between the headtracker and the 3D audio- and positioning system. The R2999M, Memsic accelerometer was used for frontal elevation data, see section 3.2.3.

3.2.3 Accelerometer - R2999M, Memsic

R2999M is an accelerometer that uses movement of internally stored gas to sense accelerations [Memsic, 2006]. It is able to detect both static (gravity) and dynamic (vibration) accelerations in two axis, X and Y. Two accelerometers combined are able to sense accelerations in three axis (XYZ) see figure 11.

![Figure 11: Two aligned Memsic accelerometers.](image)

In this application it was important that the accelerometer was insensitive to vibrations, therefore two external low pass filter was built, for x and y output respectively. The best effect for the demands in this study was achieved using following capacitors and resistors, $C = 22 \mu F, R = 100k\Omega$. For rejection of power supply noise two capacitors ($C = 22 \mu F$) and one resistor ($R = 10Ω$) were used. For a further description of the low pass filters and the power supply noise rejection see the wiring diagram in figure 10. In this study frontal and lateral acceleration data were used to calibrate the headtracker on the user’s head. Only frontal acceleration data, see figure 9, were used during the actual searching for objects in the experiment.
3.2.4 Microprocessor - BasicX-24, NetMedia

BasicX-24 is a microprocessor and is the actual “brain” of the headtracker. It gathers data from the accelerometer and compass and computes the raw data to useful information that can be sent to the 3D audio- and positioning system. Below some information describing the BasicX-24 system [NetMedia, 2006]: BasicX-24 is equipped with ROM to store the operating system, 400 bytes of RAM, 32 KBytes of EEPROM and I/O devices such as analog to digital converters (ADC). The BX-24 uses an Atmel AT90S8535 as processor. The operating system provides a multitasking environment.

Key parts of algorithms used by BasicX-24

- Gather data from compass and accelerometer.

- Compute angles from the accelerometer data: The elevation angle is input as a raw value (see figure 12) and has to be converted to angles. Raw values from angles between 0 and ±70 degrees were measured and some arcus sinus functions could be used to approximate the elevation angles without using to much processing power.

Figure 12: Raw data angle measurement procedure.

- Compass vibration compensation; the compass used was found to be a bit sensitive to vibrations caused by a human walking. This problem was solved in two steps:

  1. The headtracker was mounted upon a bike helmet with soft fabric on the inside.

  2. The headtracker does not send a compass value if it differs more than 40 degrees from the previous value. This is done in order to dismiss erroneous values that did appear when the compass was near ferrous objects or was bumped. This stopped users from turning their heads fast, as the algorithm interpreted it as an erroneous value, so when two consecutive values were interpreted as erroneous, the values were in fact interpreted as correct (head movement) instead.
• Choosing object to be found; if a button is pushed on the headtracker the headtracker sends a different object number to the 3D audio- and positioning system and consequently shifts position of the target area.

• Calibration; this function is used when the headtracker is placed on a users head and is about to send azimuth and elevation data to the 3D audio- and positioning system. When the user reports that he/she is standing in a natural pose and looking straight forward the actual values for lateral and frontal elevation are calibrated.

3.3 Audio software, AuPos

The purpose of the audio software is to play sounds in a pair of headphones. The sounds should be perceived by the listener as if they originate from a sound source outside the headphones, e.g. 3D audio. Moreover, the audio software shall also be able to be controlled by a headtracker and it shall also be able to connect to a positioning system so that the user may be tracked.

There were no software available to meet the requirements for the experiment. Therefore, a custom developed application was built in Java 1.5.

3.3.1 Development progress

The requirements were that it could play ten multiple sounds simultaneously, playback should output 44100Hz@16-bit, sound can be placed in a virtual space, HRTF-technology should create the 3D sound for headphones, an external headtracker and an external positioning system should be incorporated. Above these requirements, the audio playback latency should be 40ms. The AuPos software was developed to meet these requirements. The development was focused on a flexible system architecture because it was to be used in yet another experiment. After a month of development, the other experiment was canceled and the software was instead focused on this particular experiment and thus it was not longer a requirement that it was flexible. The requirements were changed and the software was only to be able to play one sound at a time.

At first, the idea was to implement the software in Matlab and create an interface to Java for communication with the headtracker and the positioning system. HRTF-files was downloaded from the CIPIC database ([Algazi et al., 2001]) which was already implemented in Matlab and included some test code that could be useful. Matlab is ideal for matrix computations. Soon Matlab proved bad for a real time system though and it had poor audio capabilities. Good audio capabilities was essential to the application (see section 2.2 and section 2.8). The software was developed completely in Java instead. Java had built in audio capabilities and threading which was very desirable. After a recon on the Internet concerning Java audio, the impression was that it would work well and be simple to use.

The first approach to convert wave files to Java double arrays was to use a Matlab implementation to save the Java arrays as Java objects to disk since no simple converter was found in the Java library. This was of course very inconvenient, but still it worked. By using a playback buffer in which streamed the audio array from disk, it was not necessary to load the entire sound clip to memory.

Since it was believed that it would be CPU intensive calculations for processing the audio, the pre-calculations approach was taken. It was very soon dismissed since
an original sound clip with a duration of a couple of seconds was 800MB/HRTF. That much disk space would not be available. Instead, the HRTF matrices was converted to Java double arrays and all audio processing was made in Java. At this stage, there were two possibilities to playback the sounds, either by using a thread per sound clip or by mixing all audio internally and playback the mixed audio. Since Java has an extensive and easy thread management it was simple to implement the playback as threads playing a single sound clip each.

The most CPU intensive calculation was the convolution with the HRTF. Because of the small playback buffer the calculation could only be performed on small arrays and that created an overhead in the size of the HRTF (200 samples). In this implementation, crossfades (see table 1) are used between the buffers to smooth the audio transitions from the different buffers and HRTF switches, and in that way avoiding clicks and pops. A small crossfade between each buffer was implemented and worked well; no audible pops or clicks.

It should also be mentioned that during the development of the AuPos software, there are many applications and functions developed that cannot be used since their features were to be used in another experiment which was canceled. If no such extensive and unnecessary software development had taken place, the AuPos software would probably have had a more flexible and expandable design.

### 3.3.2 Troubles during development

Two major problems were surfacing in the beginning, the audio had clicks and it did not sound exactly as the Matlab implementation. After extensive troubleshooting the problem was found in the convolution algorithm. Every buffer started 200 samples to early, resulting in a click, and the convolution algorithm was not implemented mathematically correct resulting in a deviant behavior. When corrected, the audio sounded as it should.

The CPU load had always been pretty low, about five percent on the laptop when playing a sound. It had not been a problem at all. When playing multiple sounds simultaneously for the first time the CPU was not able to play more than four sounds. This was a great problem because the software was to be used in an experiment with up to ten sound sources! After some optimizations to speed up the software, it performed only marginally higher.

An idea was that the non linearity was based on the threading so the software was rewritten to use a mixer approach and mix all audio internally before playback. It did not change the performance. This problem was however solved when the other experiment was canceled, leaving the requirement with multiple sound simultaneous. The software was instead focused on the performance in this experiment and crafted solely for one purpose. The headtracker interface was easily implemented and the Ekahau interface was also pretty simple. One downside with the Ekahau positioning system was that when the network card lost the connection to the WLAN network, the connection to the Ekahau Engine never was setup again and also that the documentation of the API was sparsely written. Once recognizing the connection problem, it was no trouble fixing it.

### 3.3.3 Design

The AuPos software is designed to be used specifically for wearable 3D audio applications. It is crafted and adjusted particularly for our experiment and experiment
equipment where we track the user. It is possible to use AuPos software with other equipment than ours with some modifications even if it is not built and designed to be used with other hardware. The software can position several wave-files in a virtual audio space using modern HRTF-technology. The user must listen via headphones for a 3D perception of the sounds. Individually made HRTF-sets can be loaded into the software to best suit the user.

When the application GUI is launched, the user first selects a file which contains the exact locations and the order of them to be found by that particular participant. The user then selects the HRTF that best match the user (see section 3.3.8). AuPos software then connects to a pre-defined Ekahau Engine to track the user position in the building. It subsequently connects to the headtracker and then plays a sound (wave-file in the application directory, see section 3.3.5) in the direction towards the object to be found.

3.3.4 Performance

The software is not optimized for CPU performance since we only use one sound at a time. The heavy convolution necessary for HRTF-calculations are CPU intensive and drains most of the CPUs resources. On the test computer the CPU load is 80% when three sounds are played simultaneous. Memory usage is mainly the .hrtf-file (∼seven megabytes, see section 3.3.5) and the .wav-file. The ChooseHRTF application has a memory load intensive phase; when it starts, it loads all .hrtf-files into memory. This may lead to memory management errors if many .hrtf-files are loaded.

The software listens to the Ekahau Engine and the headtracker and updates user position and head direction when new information is accessible. AuPos uses a 40ms playback buffer, thus minimum latency is 40ms. In reality the latency is much higher due to low headtracker update frequency and high Ekahau positioning latency.

3.3.5 File management

pos-files A .pos file is a standard ascii file with coordinates to objects, see figure 13. The first two numbers corresponds to x and y coordinates while the last number is the floor where 0 is 3rd floor, negative number is 2nd floor and positive number is 4th floor. Each row corresponds to a new object.

Figure 13: The layout of a .pos-file. The file is a plain text file and the first two numbers corresponds to x and y coordinates while the last number represents the floor.

```
298 178 -100
52 228 -100
344 309 -100
54 364 100
78 556 0
555 429 0
```

hrtf-files A hrtf-file is a Java serialized double array with size [25 azimuth][50 elevation][2 channels][200 samples] containing a complete HRTF extracted from the CIPIC database [Algazi et al., 2001].
3.3 Audio software, AuPos

.wav-files The .wav-files are standard mono wave-files with sample frequency 44100Hz and 16-bit resolution.

3.3.6 Headtracker interface

The head tracker sends data as soon as it is able to, which is approximately eight times per second, see section 3.2.1 on page 23 for details about the headtracker. The software listens to incoming data on the serial port. The head tracker device always writes four bytes at a time. The first byte is a sync byte so that the software and hardware is in sync and a correct sequence of bytes will follow. The following byte is the elevation byte. Since the elevation is between -90 and 90, we simply add 90 degrees (each byte can represent a value from 0 to 255) when the headtracker sends the data and subtracts 90 degrees when we read the data. The last two bytes are azimuth data. The first of the two bytes is either 0 or -1. In case of -1, 256 is added to the last byte data (the maximum resolution in azimuth degrees is 359). Of course, it is not the optimal solution to send this many bits only for elevation and azimuth but since the implementation was simple and fast it was a solution as good as any other. The AuPos software use a Java interface to communicate with the headtracker, thus another implementation can be implemented, e.g. a different physical headtracker.

3.3.7 Positioning system interface

The positioning system interface used in AuPos can position a user in a virtual 3D space. In the current implementation AuPos use the Ekahau Positioning Engine (see section 3.1 on page 21) to position an object. Another implementation, e.g. with another positioning system, can be implemented if desirable.

The current implementation also use the coordinates of the user to position the object in elevation. This since the system shall play a sound in \( \pm 60 \) degrees elevation when the user is on a different floor than the object (see section 4.2.2 for information on implementation details).

3.3.8 Graphical User Interface

The best HRTF fit is chosen by listening and comparing different HRTF-sets using clickable buttons that corresponds to different HRTF-sets. The user interface can be seen in figure 14. This method is based on a method by [Seeber and Fastl, 2003] described in short in section 2.4.

Figure 14: GUI for choosing the user’s best HRTF fit in a set of individual HRTFs.

3.3.9 Quality control

An iterative development process has been used when designing the software platform. No formal development method has been used, but in some extent the PSP3 software
process [Humphrey, 1996] has been used as a foundation in the development process. The design is modular and every module has been developed in an object-oriented fashion. The use of interfaces to easily switch implementation are used for the HRTF-positioning, user positioning and for the headtracker.

Each Java method has been thoroughly tested and approved using formal testing methods [Humphrey, 1996]. Every module has been tested as well as the whole system. Since this system would be used when walking, it was important to do real user tests to find problems that did not appear when running the software on a stationary setup. The network connection was becoming a problem when the connection reconnected after a disconnection. Therefore, the system was designed to survive severe errors and always make a best effort to recover when an error occurs. All errors are logged on the stdout.

Even if it is desirable, the software is not fully object-oriented. This is however not a problem since the implementation is working and the software is not supposed to be further developed.
4 Pilot test

A small scale pilot test was carried out with three participants (figure 15) to find out the possible weaknesses in our 3D audio system and potential weaknesses in the experiment method and setup. To better evaluate the results of the test, the three participants used were all supervisors to this study. Each of them were given an unique test group and conducted a single mission. The experiment setup was evaluated as well as the aid. The pilot test resulted in determining the maximum time of a mission, fifteen minutes, based on the “no aid” participant result (see section 5). Guidelines for object placement was reworked; objects may only be placed in an area visible at a few meters distance and in chest level. A more detailed version of the experiment leader guide (see section 5.4) was created so that every participant received the same information and so that the information was consistent and complete during the experimental phase.

Figure 15: Example of three participants from the pilot test.

It was suggested that a specific set of objects and locations were used during the experiment, but with a different order for each participant. The rationale was that every participant had to find the same objects. The object positions should be the same so all participants had equally hard to find the object. Using a random object order for each participant would make it possible to calculate if there are learning effects, and implicitly remove the learning effects from the resulting data set.

Trouble with the headtracker gear was found. The head gear used to fasten the headtracker to was not stable enough to bear the stress of a walking participant. It was found insufficient to withhold the calibration as the head gear moved around on the head. Also, the head gear was uncomfortable. A new head device, a rebuilt bike helmet, was suggested as a replacement.

The audio clip used was received with positive response; every test participant found it appropriate. During the 3D audio test, it was emphasized that it was hard to hear elevation. It resulted in changing the volume function (see figure 18) to indicate if the user is on the right floor. That change is believed to have increased the performance of the 3D audio system considerably.

A few questions were added to the questionnaire (Appendix A.3) to collect valuable considerations from participant experiences. Such qualitative information was believed to be helpful when designing guidelines for future audio aids.

Some improvements in the audio system software were made. Improved handling
of network disconnection reconnections and status monitoring of Ekahau interface was added. A few minor bug fixes and a GUI for choosing a specific participant’s object order.

4.1 Specification of audio system

The two most important factors in the audio system were the use of a headtracker and the sensation of a virtual 3D audio source. It was also important that the positioning inside the building had an error less than a few meters, otherwise the error would result in a “jumping” sound source if the user was near the object. The goal was to have a headtracker with an updating frequency of 25Hz [Zhou et al., 2004] and thus a software that had a latency of less than 40ms so that the user would have a smooth audio experience. The audio playback software should be based on HRTF-technology and be able to play a sound file from any direction. HRTF-technology was assumed to be the best 3D audio technology available (see section 2.4). Since spatial hearing is depending on the audio quality (see section 2.2) so it was important to play back the audio in top notch audio headphones.

Unfortunately, these requirements were not possible to meet. The custom built headtracker (see section 3.2.1) incorporated a digital compass which only had an actual update frequency below eight Hz. The compass was also sensitive to nearby magnetic fields in the building such as elevators, electrical cables and pipes which affected the compass by rotating its direction in a nonlinear fashion. The errors in the compass was however so small that the test users did not think that it would be any significant problem to the user. Moreover, a custom audio application was built in Java 5.0 and it is able to meet the 40ms latency but not entirely able to meet the requirement that any direction could be represented. Only angles in figure 16 can be represented in the audio software due to limitations in the HRTF-library used. The CIPIC HRTF-library [Algazi et al., 2001] is used because it is free, contains many HRTF sets and it is recorded in high quality.

![Figure 16: HRTF angle measurements in frontal view (a) and side view (b)](image)

The audio software play audio in 44100Hz and 16-bit to ensure high quality sound output. Sadly, it was not suitable to use headphones in the system due to magnetic
4.2 Audio choice

interference with the headtracker. Instead, a pair of earplugs was used because they did not show any magnetic interference. The audio quality was acceptable; the deep bass frequency band is an unimportant 3D audio property (see section 2.2) and the earphones provided the quality needed according to the pilot test, even if the headphones were better off.

4.2 Audio choice

4.2.1 Sound clip choice

Several sounds were tested. Considerations where taken to fulfill requirements stated in section 2.8 on page 17. Examples of sounds ranged from music loops to chain saws to toy robots to bells. A particular bell sound that sounds like a shaking tambourine mixed with some kind of arabic jingle bells was chosen in the pilot test. The bell sound duration is approximately five seconds. The bell sound has some frequency peaks but an overall fairly flat frequency curve which is important for human localization, see section 2.8.

The most important property of the white noise is the flat frequency response which indicates that all frequencies have equal sound pressure, e.g. volume, which is optimal for human localization perception (see section 2.2). In figure 17 you can see the spectrum analysis of the white noise and bell sound. Equal sound pressure for each frequency is not expected to be reached for a pleasant sounding sound. It is of course important to strive in that direction and it cannot be expected to have a more flat frequency response in the higher frequencies than the bell sound. Also, it is somewhat troublesome to analyze the bell sound frequencies over time since it is not a “constant” sound/tone and some peaks may have a small duration which may render the audio spectrum slightly false even though it probably does not have a great impact.

4.2.2 Audio playback function

The goal was to analyze the 3D audio perception rather than different audio cues. Therefore, only some basic audio cues where incorporated in the system, e.g. high-pass/lowpass filtering or pitching the sound to indicate elevation was not used. As it is pointless to incorporate a complete physical model of for example the audio volume (you cannot hear a bell if you are on another floor 40 meters from the sound source) some other volume function had to be used. After testing different volume functions, the volume cue function in figure 18 was used. As you can see, the volume is constant if the user is on the wrong floor. If the user is on the right floor, there is significant volume change. When the user is closing in on the target, the volume is rising. When the user is less than six meters from the target, the volume is peaking. This makes it easy for a user to hear if on the right floor and to hear if near the target. If the user is on the wrong floor, the sound will be heard at a ±60 degrees elevation from the object. A correct physical representation would not be suitable for the elevation, since the physical elevation angle is very small if the user is a bit away from the object and thus cannot be percepted by the user.
Figure 17: Bell sound spectrum (top) and white noise spectrum (bottom). White noise has ultimate sound characteristics for human 3D audio perception. The frequency axis is linear from 20Hz to 22050Hz.
Figure 18: Volume cue for 3D audio. The thick line represents the volume on the right floor and the dotted line represents the volume on the wrong floor.

\[ v = 1.2 - 0.95 \sqrt{\frac{d}{15}} \]

\[ v < 0.2 : v = 0.2 \]

\[ d < 6 : v = 1 \]

5 Method

The experiment took place 20060125-20060220 at the IT-University, Gothenburg. The purpose of the experiment was to investigate if 3D audio provides useful help to locate static objects indoors. To answer our hypothesis we divided the participant into three different conditions (independent variable), No aid, 3D audio and Map. The primary outcome of interest was whether the audio group would be more successful in searching for object than the group without any aid. To do so the time for finding objects was measured (dependent variable). One other interesting aspect was how efficient 3D audio would be compared to using a map which was considered to be the most efficient aid for finding objects.

There were 24 participants taking part in this experiment. The age range was 15-26 years. 11 female and 13 male took part in the experiment. The participants were randomized into three equally sized experimental groups (no aid, map and 3D audio). The no aid group was the control group in this study. The participants were told to use the given aid to look for objects placed on different floors in a building. The independent variable in this study was the experimental groups, the dependent variable the time it took to find objects. After the experiment the participants filled in a questionnaire about their feelings about the experiment as a whole and the aid they used during the
5 METHOD

5.1 Participants

There were 24 participants in this experiment. The age range was 15-26 years old, see figure 19. 13 males and 11 females took part in the experiment. The gender representation can be seen in figure 20. Taking part in the experiment was entirely voluntarily. The requirements for being a participant were the following:

- No hearing impairment
- Not familiar with the premises at the IT-University, Gothenburg.
- Not being physically impaired e.g. does not have problems to walk in stairs.

It was important in the study that the participants had not seen our experiment setup in advance which would have given them a great advantage and the results would not have been reliable. Therefore we choose not to involve students at the IT-University as participants. The participant selection was made convenient by announcing the experiment in nearby facilities and among students (see figure 21). There was also a
recruiting campaign where people were asked if they would like to participate in an experiment. This could result in a biased participant selection, e.g. the participants are chosen mainly with academic background, participants are active and eager to break new grounds and last but not least, they are not randomly chosen. This bias does not, however, change the fact that 3D audio perception is not a knowledge based on academic background but rather based on physique (see section 2.2 and section 2.7).

Figure 21: Example of a participant from each condition.

5.1.1 Test groups

The participants were divided into three equally sized test groups (map, no aid, and 3D audio). The order in which test group each participant should be placed were set before the announcement for test subjects was made e.g. participant one in test group “map”, participant two in test group “no aid”, participant three in test group “3D audio” and so on. When participants signed up for the experiment they were placed in the corresponding group to in which order they signed up. The rationales for the different conditions are:

1. To be able to make a statement about the efficiency of the 3D audio aid and the map aid it is necessary to have one group where the participants do not have any aid at all, a control group which can be used as a baseline for efficiency.

2. The 3D audio aid is the primary aid of interest in this study.

3. Visual guidance aids are commonly used and are probably more effective than other aids.

Since visual guidance aids are most common it is interesting to use one map condition to see if this is the case also in this task. Doing so enables measurements on how effective the 3D audio aid is for locating objects compared to an aid that uses vision. It is also of interest to see the level of efficiency of the 3D audio during the particular setting of this experiment.

5.2 Apparatus

Ekahau Engine 3.1 is a commercial application which is using probability to determine the physical position of WLAN cards in a network. The engine compares the signal strengths received from the WLAN-Access Points at any given point in a building and compare these data to the sampling points gathered with Ekahau Manager.
Ekahau Manager 3.1 is a commercial application used to add maps, calibration points et cetera to the Ekahau Engine. The Ekahau system uses calibration points sampled from a wireless network to track objects. In this study a total number of 348 calibration points were measured on three floors at the IT-University (95 calibration points on floor two, 116 calibration points on floor three and 137 calibration points on floor four). The estimated maximum resolution of the calibrated tracking system is about 0.1 meters and the minimum resolution is about 4.5 meters.

Ekahau Client is a commercial application that has to be installed on every computer you wish to track. After installing Ekahau Client it is possible to choose which engine you grant for positioning of the actual laptop.

Engine laptop is a Compaq Presario 2100 (S/N: CNF3281QMG) running WindowsXP with Ekahau Engine 3.1 and Ekahau Manager 3.1 installed. This laptop is used by the experiment leader during the experiment. The laptop is connected to an internet network via a network cable. The connection speed is 100 Mbit/second.

3D audio laptop A Dell Precision M60 (Model Dell LBL, P/N: 8K493A03) running WindowsXP Professional is used with 3D audio software (see section 3.3), Ekahau Client and soundcard installed. This laptop is used by the participants during the experiment. In the 3D audio program it is possible for participants to choose suitable HRTF’s. The program is calculating the laptop’s position relative to a traced target. This is possible through connection to Ekahau engine and a headtracker. The laptop position itself in a building using Ekahau client. Ekahau Client is connected to the Ekahau engine (Engine Laptop) through a WLAN card. The laptop is connected to a headtracker through the COM-port.

WLAN card A Belkin Wireless G Notebook Card (F5D7010, MAC: 001150-043776) was installed at the 3D audio laptop for wireless connection to the Ekahau engine. The maximum connection speed is 54 Mbit/second.

Headtracker The main parts of the headtracker are a microprocessor (NetMedia, BasicX-24), an accelerometer (Memsic, MXR2999E) and an electronic compass (Honeywell, HMR3300). The headtracker is placed on a rebuilt bike helmet. The headtracker sends vertical and horizontal data about the participants head angle to the 3D audio laptop via a communication cable to the COM port. The data stream is one way only.

Communication radios Two walkie-talkies (Hoffer, SP3381, S/N: 0439078740) were used during the experiment for communication between the experiment leader and experiment assistant.

Earphones One set of earphones (Sony, MDRED21LP) were used during the experiment to play 3D audio sounds to the participants in the 3D audio group.

Soundcard One Creative labs USB Soundblaster Live! 24 (S/N: M1SB049053600081M) were used to play 3D audio sounds to the participants in the 3D audio test group.
5.3 Materials

**WLAN-Access Point** Thirty-three fixed access points (Cisco Systems, Cisco Aironet 1200 Series) were used during the experiment to track the participants’ positions on different floors in the building. Ekahau engine uses these access points to calculate the most probable position of the user.

**Stopwatch** A cellphone model Nokia 3310 is used as a stopwatch. Time is recorded in seconds.

5.3 Materials

**For the participants** Consent form, different scenarios for the different experimental group (no aid, map and 3D audio), sheets of fraction numbers, form to put your signature on for movie tickets, movie tickets, questionnaire about the experiment, pencils, blank sheets of paper, maps over the different floors of the IT-University, sheet for choosing appropriate HRTF file, heavy book, a light block of polystyrene, backpack, entry card and numbered photos of famous people.

**For the experiment leader** Step by step procedure for the experiment, sheet for in which order the participants are to seek for hidden objects, a brief introduction about 3D Audio and a pencil.

**For the experiment assistant** Pencil, observation form, entry card, numbered photos of famous people, maps over the different floors of the IT-University and sticky tape.

5.4 Design

This was a between groups design with one “no aid” condition, one “map” condition, and one “3D audio” condition. The no aid condition consisted of searching for photos of people, one at a time, without any help given concerning where the photos could be placed. The map condition consisted of searching for photos of people, one at a time, by looking at a map where a corresponding number of the actual person was added plus some irrelevant numbers. The 3D audio condition consisted of searching for photos of people, one at a time, by following a sound presented to the participant via earphones. The independent variable was in which condition the participant conducted the test, e.g. no aid, map or 3D audio. The outcome variable of interest, the dependent variable, was whether the time it took to find the photos would differ between the conditions or not.

The test groups are independent variables; all participants are randomly selected to each group and all groups do the same procedure, apart from using different aids and their setup. Time is used as a dependent variable because it is directly proportional to the efficiency in finding the object, e.g. if time is short the efficiency is high. Other variables, like distance travelled, was considered but rejected since it is hard to measure and the study would give a more useful result if time was used.

Further more, the setting used in the study, e.g. IT-universitetet, is an independent variable in such way that it may influence the way participants search. Number of rooms, total search area and the actual architectural style may impose difficulties and those difficulties are hard, if not impossible, to measure. Moreover, the actual position of the object in a given location may highly contribute to the difficulty, therefore the objects in this study was placed in a consistent manner (see figure 22) and visible within a few meters distance. The actual images used on the cards are high contrast.
black and white images with one single purpose, to make the experiment more real (see appendix A.1 for the user scenarios and figure 23 to see the cards). The size of the images is A5 which is easy to locate from a few meters distance.

5.4.1 Minimizing biases

To treat each participant in the same way a number of actions were taken. The experimental leader and the experimental assistant never switched functions during the tests. By using an experiment leader process script, it was ensured that the information given to the participant was the same to all participants. Each participant was offered carbonated water to feel comfortable. The experimental assistant followed each participant during the tests and took notes about circumstances that could influence the result e.g. a crowd of people at a certain location et cetera. The experimental assistant was always ready to help the participant if he could not open a door et cetera.

The participants was to have comparable equipment mounted (3D audio group used

Figure 22: Placement of all card objects. The area around the card object is highlighted. Each card object is of size A5.
Figure 23: Cards to be found are images of Victoria Bernadotte (06), Jesus (12), Carl-Philip Bernadotte (14), Carl XVI Gustaf Bernadotte (15), Madeleine Bernadotte (22) and Hulk Hogan (28). Each card object is of size A5.

a laptop and other groups used a book to simulate the weight of the laptop to save the laptop for unnecessary stress and to reload its batteries) so that the participants did not differ from each other. To minimize outer factors, e.g. inevitable biases, such as number of students currently in the building, student breaks or refurnishing, the experiment conditions were spread out in time during the whole study so an unexpected change would influence all test groups as equally as possible. To minimize learning effects, the mission order in each experiment were randomized. Every participant practiced using their aid during a trial mission. The starting point of the experiment was kept constant for all participants to avoid advantages such as seeing an object directly when the experiment start.

5.5 Procedure

The participant was greeted and given a brief presentation about 3D audio and its current applications and benefits. The participant was then asked to read the consent form (Appendix A.4) and fill it out appropriately. A summary in swedish was held for those who wanted. The experiment leader informed the participant that the participant was
not the subject of interest, but rather the 3D audio system, and pointed out that the participant was participating anonymous. Then the participant was asked to read the test scenario according to the participant’s test group.

If the participant was in the 3D audio test group, a matching HRTF was chosen with the ChooseHRTF-application shown in section 3.3.8.

The helmet and backpack was mounted on the participant. If the participant was in the 3D audio test group, the laptop was placed in the backpack, otherwise a book was used. The headtracker was fastened and calibrated on top of the helmet, see figure 24.

The experiment assistant hide the test card object 22 (figure 23), also referred to as “the person” or “the object”, on location 22 on the map (figure 25).

The participant was informed that the first mission is a trial mission where the participant may ask any question to the experiment leader or the experiment assistant, and where the result will not be used in the study. The participant was also told that no questions would be answered after the trial mission. The participant was then given an entry card to the building to open the doors in the stairways. The participant was informed that the test card object was to be found on the current floor, the floor above or the floor below. Moreover, the participant was informed that the objects did not reside in any enclosed areas, e.g behind a door, that they were visible from a few meters distance and that they were placed in approximately chest level, see figure 22. The participant was told that the goal was to find the objects as fast as possible, using the aid (map, 3D audio or no aid) as guidance, and bring the objects back to the experiment leader. The participant was instructed to walk in a moderate pace, not using the elevators, not talking to anyone. It was pointed out that the aid was only to be used as guidance and that the participant was expected to look around in the surroundings. A total of five objects, plus the test object, was told to be found - all in separate missions, see figure 22.

The participant was given a test card object and was then instructed to fetch the copy
5.5 Procedure

Figure 25: Map of the IT-University, Gothenburg, where the experiment was conducted. Floor 2 (left), floor 3 (right) and floor 4 (bottom).

of that object which were hidden in the building. The stop-watch was started when the participant received the object. The experiment assistant followed the participant in a discreet manner and made notes while the participant was searching for the object. When the object was found the stop-watch was stopped and the object, participant and experiment assistant returned to the experiment leader.

If the participant had a map, it was returned to the experiment leader. The participant was asked for any further questions and when all questions were answered the participant was asked to fill out the fraction paper (appendix A.2). The participant was not aware of the fact that the fraction paper only served as a distraction when the experiment assistant hide the next object. The door to the experiment room was always closed when the participant was inside it and the experiment assistant made effort to keep the time for hiding objects constant, so that the participant could not calculate approximate distance to the hidden object in respect to the experiment assistant’s time away. When the experiment assistant returned, the participant was asked if the participant was ready to start next mission. If the participant was in the 3D audio test group, the next object to be found was chosen with the “next object” button placed on the headtracker. A new test card object was handed out to the participant. If in the map condition, the participant received the map together with the test card object. The
stop-watch was started when the participant received the item(s).

The experiment assistant was following the participant in a discreet fashion and made notes while the participant was searching for the object. During the missions, the experiment assistant was quiet and did only speak in ambiguous situations that did not jeopardize the experiment design. When the object was found, the stop-watch was stopped and the object and participant returned to the experiment leader. If the object was not found during a time period of fifteen minutes, the mission was canceled and the experiment assistant lead the participant to the object and then returned to the experiment leader. Fifteen minutes was recorded in the observation protocol as well as a note indicating that the time limit had been reached.

When all five missions were accomplished, the entry card was returned to the experiment leader and the equipment was unmounted from the participant. Then the participant filled out a questionnaire about the experiment. For participant data and the participant’s experience of the experiment see section 5.5.1. When finished, the participant signed up for a movie ticket, was thanked for the participance and was then escorted out of the building.

5.5.1 Questionnaire

Upon successful completion of each session, the subject was asked to fill out a questionnaire. The questionnaire is in Swedish and can be seen in appendix A.3. In section A the participant filled in sex, age and test group. That information is used only to individualize the participant and to have a record of the sex ratio and age span. In question four, the participant is asked to give a summary of the experiment implementation to see how the participant was relating to the experiment setup and the over all opinion about the experiment. It would give a pretty good indication if the experiment was professionally conducted and thus if the results are reliable. If the participant was in the 'no aid' group, section B and C were not to be filled out.

In section B, the aid was evaluated. In question five and six the participant was asked if the participant found the aid helpful when searching for an object and if the participant would use a similar aid if in a real life job situation. These questions are interesting because lack of interest and lack of confidence in the aid will reflect an unwillingness to use such an aid and thus the aid is obsolete or has to be re-worked.

Question seven and eight reflect how easy the aid is to use and if it acquires a learning time. This information can be used together with statistical data to see if the results improved over time. It is of course a bit misleading when the map aid is a commonly used and widely accepted aid that all participants most certainly have used before. Even so, it is of interest to evaluate the 3D audio aid to gain knowledge of learning effects.

In questions nine through eleven, no predefined answers were given and the participant were to give qualitative statements for improvements of the aid. These answers may be used as guidelines to what can be experienced as annoying and what users treasure for future development.

Question twelve reflects the overall focus on the task necessary for using the aid. By evaluating the focus level, guidelines can be worked out to elucidate if the aid is suitable in environments already accompanied by heavy mental workload.

Question thirteen was only to be answered by the 3D audio participants. It reflects the importance of several factors in 3D audio such as latency and choice of audio clip. The results can be used as guidelines when designing a 3D audio system to stress the
different factors and their importance to the users overall experience of the system. The results can be seen in figure 28 and figure 27.
6 Results

6.1 Experiment results

To be able to draw statistically valid conclusions from the data sampled during the experiment SPSS 13.0 was used to perform an ANOVA and a Post-Hoc analysis. For an explanation of what Post-Hoc and ANOVA means, see below:

- ANOVA: “The One-Way ANOVA procedure produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. Analysis of variance is used to test the hypothesis that several means are equal” (SPSS 13.0 for Windows)

- Post-Hoc analysis: “In addition to determining that differences exist among the means, you may want to know which means differ. There are two types of tests for comparing means: a priori contrasts and post hoc tests. Contrasts are tests set up before running the experiment, and post hoc tests are run after the experiment has been conducted. You can also test for trends across categories” (SPSS 13.0 for Windows)

A one-way analysis of variance (ANOVA) was run on the data from the study. The overall significance level was set at 5%. The effect of condition was found to be significant $F(2, 117) = 30.61$, ***$P < .001$ e.g. the time it took to find objects depended on in which condition the participant did the test, see table 2. The variances was not equal between the groups. No learning effects, sex or age differences were found in any of the conditions.

Table 2: An ANOVA analysis of variance.

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1941771.7</td>
<td>2</td>
<td>970885.858</td>
<td>30.608</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3711286.9</td>
<td>117</td>
<td>31720.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5853058.6</td>
<td>119</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean time (M) and standard deviation (SD) to find the objects were the following: No Aid (M = 384.25, SD = 262.176), 3D Audio (M = 149.98, SD = 156.03), Map (M = 89.20, SD = 45.61). For a box plot of the results see figure 26. Post-Hoc analysis (TAMHANE), table 3, revealed significant differences between the 3D audio condition and the no aid condition ($***P < .001$), and between the map condition and the no aid condition ($***P < .001$). 3D Audio was significantly more efficient than using no aid, also map was significantly more efficient than using no aid. However there was no significant difference between the 3D audio condition and the map condition ($P = .066$). The overall significance level for all post-hoc tests was set at 5%.

6.1.1 Experimental assistant protocols

- Two participants claimed that they missed an object because they did not dare to make a closer look due to the people in the area, e.g. two objects out of 120 was

---

One star (*) means that the result is not due to chance with 95% confidence. P is less or equal with .05. Two stars (**) means that the result is not due to chance with 99% confidence. P is less or equal with .01. Three stars (***) means that the result is not due to chance with 99.9% confidence. P is less or equal with .001.
6.2 Questionnaire results

The questionnaire data does not differ between the map and the 3D audio other than that the map was experienced very easy to use and the 3D audio was only experienced fairly easy to use. The participants using the 3D audio aid rated the relative direction (use of headtracker) and the 3D perception as very important factors. The positioning affected by the current situation in the building.

- For three participants in the no aid group there were ongoing lectures on places where participants usually could search in, e.g. the search space was diminished.

Figure 26: A box plot of the three conditions: no aid (1), 3D audio (2) and map (3).

Table 3: A Tamhane post-hoc analysis with time as dependent variable.

<table>
<thead>
<tr>
<th>(I) Condition</th>
<th>(J) Condition</th>
<th>Mean Difference (I-J)</th>
<th>Std Error</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>234.275*</td>
<td>48.239</td>
<td>.000</td>
<td>115.98 - 352.57</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-205.050</td>
<td>42.076</td>
<td>.000</td>
<td>190.35 - 399.75</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-343.275*</td>
<td>48.239</td>
<td>.000</td>
<td>-352.57 - 115.98</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>69.775</td>
<td>25.703</td>
<td>.066</td>
<td>-2.93 - 124.48</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-265.050*</td>
<td>42.076</td>
<td>.000</td>
<td>-399.75 - 190.35</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-69.775</td>
<td>25.703</td>
<td>.066</td>
<td>-124.48 - 2.93</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.
accuracy was experienced less important, and the system latency and the sound played was neither important nor unimportant. The questionnaires point out that many 3D audio users experience that their mental focus on the audio is decreasing during the experiment and that they “learn to hear” over time. Such effects cannot be read from the data set though. One user in the 3D audio condition complained about the lag introduced by the headtracker.

Most users rated the overall conduction of the experiment as very well carried out.

Figure 27: Questionnaire results for 3D audio and map conditions. Lower is better. See appendix A.3.

![Figure 27: Questionnaire results for 3D audio and map conditions. Lower is better. See appendix A.3.](image)

Figure 28: Questionnaire results for 3D audio specific questions. Lower is more important. See appendix A.3.

![Figure 28: Questionnaire results for 3D audio specific questions. Lower is more important. See appendix A.3.](image)
6.3 Comments from participants

Some comments are answers from experiment assistant questions and some are spontaneous comments.

6.3.1 3D audio condition

“Lättare att höra uppåt än nedåt” - Participant 6

“Svårt att se sig om samtidigt som man lyssnar” - Participant 6

“Det är ju bara att följa ljudet” - Participant 6

“Svårt att höra rätt våning” - Participant 9

“Jag hörde rätt våning. Det var tydlig ljudskillnad” - Participant 9

“Svårt med rätt våning” - Participant 12

“Man blir ganska förvirrad, svårt att höra var...” - Participant 12

“Det låter som om det kommer underifrån” - Participant 15

“(Ljudet) låt lika dant en bit ifrån (~6m) som vid Hogan. Gjorde mig förvirrad” - Participant 15

“Lättare när man vant sig” - Participant 15

“Hörde elevation direkt” - Participant 15

“Viktigt att höra rätt våning” - Participant 15

“Var så stark volym att riktningen var svår att höra. Hördes direkt att jag gått åt fel håll så jag gick runt för jag tyckte det skulle ta lika lång tid som att gå tillbaka” - Participant 15

“Elevation hördes väldigt tydligt” - Participant 15

“Jag vet ungefär vilken volym det ska vara och hörde direkt att 3 var fel våning. Tyckte det låt lite ovanför så jag chansade uppfåt först” - Participant 17

“Känns som jag fått upp mer fart nu” - Participant 17

“Jag tycker de flesta låter bakifrån” - Participant 17 about the HRTF-selection

“Det här var skitsvärt” - Participant 21

“Svårt på samma våning när man börjar med starkt ljud” - Participant 21

“Jag är seg i huvudet efter bråktalen” - Participant 21

“Har lärt mig nu” - Participant 21

“Ljudet hoppar fastän jag inte rör mig” - Participant 24

“Kändes neråt” - Participant 24

“Hade det inte suttit folk hade jag sett den direkt” - Participant 24
6.3.2 Map condition

“Blev lättare på slutet för då visste jag vad jag skulle titta efter” - Participant 18

“Svårare att orientera i skogen” - Participant 13

“Jobbigt att folk tittar på mig” - Participant 1

6.3.3 No aid condition

“Ska man ändra logiken i hur man letar?” - Participant 23

6.4 Experiment assistant protocols

“Det verkar som en del pseudoeffekter ”lurar” deltagarna, t.ex att det låter svagare utan att det gör det. Bra att det inte vet hur systemet är uppbyggt.”

There were notes that some participants walked slower and some walked faster and that they were ordered to walk in a moderate pace. One 3D audio participant always walked to the right floor on the first try. Most other comments were about the order of the floors searched and how the participants seemed to experience the situation. That information is excluded in this report because it was found to be irrelevant to this study.
7 Discussion

7.1 Evaluating method

The experiment setup was very carefully chosen to minimize biases in order to gain reliable results. During the tests, everything was disciplined and anything out of the ordinary was recorded by the experiment assistant. The software logged all errors so troubles could be traced back to software failures. The experiment was successful and no errors could be traced back to the software or hardware, which suggest that the reliability of the experiment and the whole study is high.

The procedure of searching for objects inside a building with other people in it might be a controversial setup to some, but in this study no major side-effects from it has been prominent. Only two participants blamed people (did not dare to look to close) for having trouble with finding an object and that is not to be considered a problem.

One small drawback with the method was the randomization of all objects for all participants. It would have been better to randomize eight entities so each conditions had exact the same eight object orders. In that way, no condition could have a “harder” object order, if such even exist. The process of randomization will however give a guarantee that the groups are equal in a statistical fashion which of course is the most important factor.

Some participants were disturbed by ongoing lectures during the search. The participants were however instructed that the object could not be found in that areas thus making the search area smaller. Since these areas were approximately 10x15 meters, which is only 3% of the total area, and only the no aid condition was affected (which was significantly slower than the other conditions) no actions were taken to compensate the results.

The use of a static map does make the map condition a bit harder compared compared to using a dynamic map. A dynamic map with a pointer that points in the direction of the object would have been a more fair comparison, e.g. it would have given the user the same information as the audio aid but in a visual mode. The use of a static map does make sense though; it is a common aid for searching.

The environment where the test is conducted is crucial for the result. A labyrinth would render the audio aid useless while a complex and symmetric building would increase the difficulties to navigate with a map. It ought to be much harder for the condition with no aids if the objects were concealed.

7.2 Evaluating experiment

This study concludes two major results, namely that using 3D audio is significantly better than using no aid at all and that there are no significant differences between using a map and using 3D audio (see table 3). It means that 3D audio is as good as a map when a user shall try and find an object in an unknown environment. This was a surprising result that hopefully will open up for new audio research and applications.

7.2.1 Performance

A main characteristic in the 3D audio condition is that the participants have trouble hearing elevation, but heard azimuth without problems. This resulted in the participants searching on the wrong floor but “at the right place” on some occasions but mostly
spending rather much time in the stairways listening to the volume cue. Doing this study in a one-floor environment, the 3D audio would most probably beat the map.

It should noticeable that the 3D audio condition has a larger standard deviation and a higher median than the map condition, see figure 26. Another noticeable feature is that the lowest quartile is about the same size for both 3D audio and map condition. This may indicate that 3D audio is a more “unstable” aid in the sense that it might sometimes take longer time to find an object compared to the map aid which shows little deviation in time.

7.2.2 Comments from participants

Many comments regarding the 3D audio is that it is hard to hear elevation. Only one 3D audio user heard the elevation clearly and never went to the wrong floor. Some comments are about how the participant experienced that they learned to hear the 3D audio cues over time. One participant found it hard to use visual input along with the audio input. Overall, the comments from the participants can conclude the idea that elevation presentation is the bottle-neck of the 3D audio system.

7.3 Evaluating questionnaire

The results of the questionnaires suggest that a 3D audio presentation system should focus on the 3D experience and the headtracker. The participants considered them as the most important features of a 3D audio system. The sound played was not highly rated by the participants, but is in fact a key element in the 3D audio experience (see section 2.8) which was rated very important and thus should be considered in a system design.

Other than that, the questionnaires only concluded that the experiment was well conducted and that the 3D audio system was working fine.

7.4 Evaluating 3D audio implementation

Most users in the 3D audio condition had troubles with detecting the elevation of the sound but had no problems detecting azimuth. In the map condition, it was the other way around; the map displayed unambiguous elevation information, e.g. the floor number. Users in the map group had however some minor troubles to detect the orientation and direction to the object.

Most users of the 3D audio system rated the headtracker and the 3D experience as key elements and did not think the sound clip choice mattered for their performance. Only one user complained about the headtracker refreshing rate, and thus it is probable that a slow update frequency of a headtracker does not imply that the display is bad. This means that the accuracy of the spatial positioning is not as important either since the user is only guided towards the object.

7.5 Future work

7.5.1 Audio cues

This study recommends different audio cues to represent different information in the audio domain. As recommended by participants and as could be seen in the experiment, the elevation information in pure 3D audio is hard to grasp. When navigating inside
buildings, it would probably be a good idea to have yet another cue for elevation. Pitching the sound playback to indicate if the object is on another floor might be a good idea.

Also, there are several ways to display the information. How many sounds can a human hear simultaneous yet interpret the correct spatial position? If there are too many objects simultaneous, how should they be displayed to decrease mental workload? Playing all sounds simultaneous would result in the user using the “cocktail party” effect (see section 2.2) to distinguish a particular sound. That will probably work if the user is only interesting in one object at a time. Another way is to display the sounds in a similar way as the traditional radar screens, e.g. sweeping the azimuth and play the sounds in radar focus more intense. That would decrease the number of simultaneous sounds but it would introduce latency in the magnitude of the radar sweep frequency. Yet another way is to have a “focus cone”, e.g. to display the sounds in the head direction more intense. That would make the display more usable when wanting to focus on a particular spatial position. The downside would be that it might not be as useful for a warning system which is dependable of the user perception in full 3D to acknowledge all warnings.

How should the sounds be chosen for maximum simultaneous presentation? Auditory icons are sounds that represent single object or events; a bicycle bell represent a bike and a bark represent a dog. Such sounds might not be ideal for 3D audio since they might not have the optimal spectral features (see section 2.8). The sounds should also have some sort of orthogonal property to not obfuscate other sounds in near spatial areas. There are still much research in that field, especially to create usable sounds and not only guidelines.

7.5.2 Headtracker

For outdoor- or indoor tasks similar to the search task described in this report, this study suggests a somewhat different solution of a headtracker than the one developed for this study. To build a more reliable headtracker regarding azimuth data, this study suggests a combination of an electronic compass and a gyroscope (with higher update frequency), e.g. recalibrate the gyro using the compass every thirty seconds or so. This combination will give fast communication, insensitiveness to drift, insensitiveness to magnetic fields and insensitiveness to vibrations and other outer interferences.

7.6 Suggested applications

7.6.1 Working with 3D audio guidance

The 3D audio technology might be used by several professions that require other input than visual. In a smoke-filled building, a visual display may not be the most suitable choice for navigation presentation. For firefighters performing rescue work in a burning building, it is more important to know the relative position to the target than the current position. Moreover, soldiers that cannot take their visual focus from the battle field yet must be able to distinguish between friendly and hostile troops to avoid engaging in friendly fire. Thus, a 3D audio presentation is very suitable to such tasks. Other groups might include persons with physical impairments that can be guided by their ears rather than their eyes. Finding packages in big warehouses et cetera where the system is automatically updated when a new purchase order has arrived and direct forklift trucks to the location of the package(s). It could simply be a presentation system that displays
warning signals in work situations where reaction times are crucial, e.g. pilots and certain control board personnel.

7.6.2 The development of an adaptive 3D audio system

- Wrist input unit (choose target area, target person, HRTF). Communication via Bluetooth with CPU. The unit also measures the owners’ physical condition (heart rate, blood pressure, body position) and sends an alarm via the CPU to the colleges if the condition is critical. The colleges can hereby track the injured person.

- Headphones that add the 3D audio information to the surrounding sounds e.g. an extra artificial layer of the soundscape. Peltor for instance has developed hearing protectors (ComTac MT15H68FB, Peltor), with active volume function that can be used. By using such techniques you are able to add extra information (artificial 3D audio) to the surrounding sounds, and not like in the experiment described in this report, somewhat block the surrounding sounds with 3D audio.

- CPU in protected case on the back. CPU receives input from wrist unit via Bluetooth, plays 3D audio sounds to the user, and keeps track of the user and the targeted area or targeted person(s) the user wish to track. The CPU automatically switch to the most accurate positioning system at the moment (WLAN, UWB, GPS), alternatively combine different systems if, for example the user is outdoor and the target is indoor.

- The input unit and the CPU may be updated with new applications, maps, calibrations et cetera via Bluetooth from an external computer (PDA, Laptop or PC).
References


REFERENCES


A Appendix

A.1 User scenarios

The participant read the user scenario corresponding to the participants condition to make the experiment more real to the participant.

Figure 29: One user scenario for each condition. The text is in swedish.

Experimentscenario, 3D-Audio
Du är anställd på ett större sjukhus i Mellansverige. Avdelningen du jobbar på är specialiserad på vård av förvirrade personer. En stor del av tiden får du ägna åt att leta reda på patienter som irrar runt på avdelningen. Sjukhuset har på grund av detta införskaffat ett positioneringssystem som bland annat medför att du kan höra var en viss patient befinner sig.

Uppgift:

Experimentscenario, Utan hjälpmedel

Uppgift:

Experimentscenario, Karta
Du är anställd på ett större sjukhus i Mellansverige. Avdelningen du jobbar på är specialiserad på vård av förvirrade personer. En stor del av tiden får du ägna åt att leta reda på patienter som irrar runt på avdelningen. Sjukhuset har på grund av detta införskaffat ett positioneringssystem som bland annat medför att du kan göra kartutskrifter med en viss patients position markerad på kartan.

Uppgift:
A.2 Fraction papers

During the time when the experiment assistant hid the next object, the participant was
engaged in calculating fractions as distraction. The task was to mark the greatest of
each fraction pair and to complete as many fraction pairs as possible.

Figure 30: Fraction calculations to distract the participant between tasks.

<table>
<thead>
<tr>
<th>Fraction Pair</th>
<th>Result</th>
<th>Fraction Pair</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/6</td>
<td>1/2</td>
<td>3/4</td>
<td>1/2</td>
</tr>
<tr>
<td>2/3</td>
<td>1/2</td>
<td>4/5</td>
<td>2/3</td>
</tr>
<tr>
<td>3/5</td>
<td>1/2</td>
<td>5/6</td>
<td>1/2</td>
</tr>
<tr>
<td>4/5</td>
<td>1/2</td>
<td>6/7</td>
<td>1/2</td>
</tr>
<tr>
<td>5/6</td>
<td>1/2</td>
<td>7/8</td>
<td>1/2</td>
</tr>
<tr>
<td>6/7</td>
<td>1/2</td>
<td>8/9</td>
<td>1/2</td>
</tr>
<tr>
<td>7/8</td>
<td>1/2</td>
<td>9/10</td>
<td>1/2</td>
</tr>
<tr>
<td>8/9</td>
<td>1/2</td>
<td>10/11</td>
<td>1/2</td>
</tr>
<tr>
<td>9/10</td>
<td>1/2</td>
<td>11/12</td>
<td>1/2</td>
</tr>
<tr>
<td>10/11</td>
<td>1/2</td>
<td>12/13</td>
<td>1/2</td>
</tr>
<tr>
<td>11/12</td>
<td>1/2</td>
<td>13/14</td>
<td>1/2</td>
</tr>
<tr>
<td>12/13</td>
<td>1/2</td>
<td>14/15</td>
<td>1/2</td>
</tr>
<tr>
<td>13/14</td>
<td>1/2</td>
<td>15/16</td>
<td>1/2</td>
</tr>
<tr>
<td>14/15</td>
<td>1/2</td>
<td>16/17</td>
<td>1/2</td>
</tr>
<tr>
<td>15/16</td>
<td>1/2</td>
<td>17/18</td>
<td>1/2</td>
</tr>
<tr>
<td>16/17</td>
<td>1/2</td>
<td>18/19</td>
<td>1/2</td>
</tr>
<tr>
<td>17/18</td>
<td>1/2</td>
<td>19/20</td>
<td>1/2</td>
</tr>
<tr>
<td>18/19</td>
<td>1/2</td>
<td>20/21</td>
<td>1/2</td>
</tr>
<tr>
<td>19/20</td>
<td>1/2</td>
<td>21/22</td>
<td>1/2</td>
</tr>
<tr>
<td>20/21</td>
<td>1/2</td>
<td>22/23</td>
<td>1/2</td>
</tr>
<tr>
<td>21/22</td>
<td>1/2</td>
<td>23/24</td>
<td>1/2</td>
</tr>
<tr>
<td>22/23</td>
<td>1/2</td>
<td>24/25</td>
<td>1/2</td>
</tr>
<tr>
<td>23/24</td>
<td>1/2</td>
<td>25/26</td>
<td>1/2</td>
</tr>
<tr>
<td>24/25</td>
<td>1/2</td>
<td>26/27</td>
<td>1/2</td>
</tr>
<tr>
<td>25/26</td>
<td>1/2</td>
<td>27/28</td>
<td>1/2</td>
</tr>
<tr>
<td>26/27</td>
<td>1/2</td>
<td>28/29</td>
<td>1/2</td>
</tr>
<tr>
<td>27/28</td>
<td>1/2</td>
<td>29/30</td>
<td>1/2</td>
</tr>
<tr>
<td>28/29</td>
<td>1/2</td>
<td>30/31</td>
<td>1/2</td>
</tr>
<tr>
<td>29/30</td>
<td>1/2</td>
<td>31/32</td>
<td>1/2</td>
</tr>
<tr>
<td>30/31</td>
<td>1/2</td>
<td>32/33</td>
<td>1/2</td>
</tr>
<tr>
<td>31/32</td>
<td>1/2</td>
<td>33/34</td>
<td>1/2</td>
</tr>
<tr>
<td>32/33</td>
<td>1/2</td>
<td>34/35</td>
<td>1/2</td>
</tr>
<tr>
<td>33/34</td>
<td>1/2</td>
<td>35/36</td>
<td>1/2</td>
</tr>
<tr>
<td>34/35</td>
<td>1/2</td>
<td>36/37</td>
<td>1/2</td>
</tr>
<tr>
<td>35/36</td>
<td>1/2</td>
<td>37/38</td>
<td>1/2</td>
</tr>
<tr>
<td>36/37</td>
<td>1/2</td>
<td>38/39</td>
<td>1/2</td>
</tr>
<tr>
<td>37/38</td>
<td>1/2</td>
<td>39/40</td>
<td>1/2</td>
</tr>
<tr>
<td>38/39</td>
<td>1/2</td>
<td>40/41</td>
<td>1/2</td>
</tr>
<tr>
<td>39/40</td>
<td>1/2</td>
<td>41/42</td>
<td>1/2</td>
</tr>
<tr>
<td>40/41</td>
<td>1/2</td>
<td>42/43</td>
<td>1/2</td>
</tr>
<tr>
<td>41/42</td>
<td>1/2</td>
<td>43/44</td>
<td>1/2</td>
</tr>
<tr>
<td>42/43</td>
<td>1/2</td>
<td>44/45</td>
<td>1/2</td>
</tr>
<tr>
<td>43/44</td>
<td>1/2</td>
<td>45/46</td>
<td>1/2</td>
</tr>
<tr>
<td>44/45</td>
<td>1/2</td>
<td>46/47</td>
<td>1/2</td>
</tr>
<tr>
<td>45/46</td>
<td>1/2</td>
<td>47/48</td>
<td>1/2</td>
</tr>
<tr>
<td>46/47</td>
<td>1/2</td>
<td>48/49</td>
<td>1/2</td>
</tr>
<tr>
<td>47/48</td>
<td>1/2</td>
<td>48/49</td>
<td>1/2</td>
</tr>
<tr>
<td>48/49</td>
<td>1/2</td>
<td>49/50</td>
<td>1/2</td>
</tr>
<tr>
<td>49/50</td>
<td>1/2</td>
<td>50/51</td>
<td>1/2</td>
</tr>
<tr>
<td>50/51</td>
<td>1/2</td>
<td>51/52</td>
<td>1/2</td>
</tr>
<tr>
<td>51/52</td>
<td>1/2</td>
<td>52/53</td>
<td>1/2</td>
</tr>
<tr>
<td>52/53</td>
<td>1/2</td>
<td>53/54</td>
<td>1/2</td>
</tr>
<tr>
<td>53/54</td>
<td>1/2</td>
<td>54/55</td>
<td>1/2</td>
</tr>
<tr>
<td>54/55</td>
<td>1/2</td>
<td>55/56</td>
<td>1/2</td>
</tr>
<tr>
<td>55/56</td>
<td>1/2</td>
<td>56/57</td>
<td>1/2</td>
</tr>
<tr>
<td>56/57</td>
<td>1/2</td>
<td>57/58</td>
<td>1/2</td>
</tr>
<tr>
<td>57/58</td>
<td>1/2</td>
<td>58/59</td>
<td>1/2</td>
</tr>
<tr>
<td>58/59</td>
<td>1/2</td>
<td>59/60</td>
<td>1/2</td>
</tr>
<tr>
<td>59/60</td>
<td>1/2</td>
<td>60/61</td>
<td>1/2</td>
</tr>
</tbody>
</table>

A.2 Fraction papers
A.3 Questionnaire

After each completed experiment, the participant filled out the questionnaire accordingly. The questionnaire is mainly about the participant’s experience of the aid used.

Figure 31: Questionnaire to be filled out by the participant after the experiment.
Before the experiment began, the participant signed the consent form which basically says that the participant will be anonymous and that the participant’s results can be used in the study.

Figure 32: Consent form.
A.5 Participant object order

The complete list of all participants object order.

Figure 33: List of each participant’s object order.
A.6 HRTF selection instructions

For choosing the optimal HRTF from the HRTF-set, these translated instructions from [Seeber and Fastl, 2003] was used.

Figure 34: HRTF selection instructions in swedish.

Välj ut vilket ljud som passar dig bäst

Steg 1

- Muschlicka på de olika siffrorna (1 till 11) för att lyssna på olika presentationer av bjällror
- Skriv ner de fem nummer som bäst stämmer in på följande beskrivning:
  Ljudet ska låta som att det kommer från en ljudkälla en bit bort från kroppen, framför dig. Ljudet ska röra sig fram och tillbaks framför dig.

Steg 2

- Nu får du endast valja att lyssna på de ljud du valde ut i steg 1.
- Du ska nu välja ut det ljud som bäst motsvarar följande kriterier.
  - Ljudet uppfattas röra sig mellan -40 grader till +40 grader relativt den riktning du är vänd mot (TIPS: -90 grader respektive + 90 grader är rakt ut åt sidorna)
  - Ljudet rör sig vågrätt i lika stora steg
  - Ljudet respektive som om det rör sig i höjdled
  - Ljudet uppfattas som att det rör sig framför dig hela tiden
    - på samma avstånd
    - och helst långt borta (ca en meter)