USING 3D AUDIO GUIDANCE TO LOCATE INDOOR STATIC OBJECTS

Samuel Sandberg
Email: kv00ssg@cs.umu.se
Chalmers University of Technology

Calle Håkansson
Email: e0calle@etek.chalmers.se
Chalmers University of Technology

Niklas Elmqvist, Philippas Tsigas and Fang Chen
Email: {elm,tsigas,fanch}@cs.chalmers.se
Chalmers University of Technology

Is 3D audio an interesting technology for displaying navigational information in an indoor environment? This study found no significant differences between map- and 3D audio navigation.
The user tasks tested involved finding objects in a real office 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 earphones. 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 study also suggests that special cues should be added to the pure spatial information to emphasize important information.

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

3D audio technology has improved considerably in recent years and can these days generate high quality sounds with accurate positions in a full three-dimensional space. More and more applications with 3D sounds can be found in, for example, the music industry, visualization applications and computer games. Three dimensional audio signals can be used in many ways, such as shifting the attention of the user (warning signals), aiding speech recognition of multiple simultaneous sources (the "cocktail party" effect), for rendering a virtual "audio map", etc.

The 3D audio effect is very subtle and thus requires high quality audio with a carefully chosen audio feed. There are several ways to achieve the 3D effect via headphones; the most delicate way is to process the audio through individualized Head Related Transfer Function (HRTF) filters that mimic the user's ears, pinnae, torso and head (King and Oldfield, 1997; Chen and Carlander, 2003). The HRTF filters are basically impulse responses of two particular ears sampled from different distances and different angles. Previous research on the spatial precision of the human audio system, e.g. the ability to determine the location of a sound source (Walker and Lindsay, 2003; Murphy, Kelly, and Tew, 2002; Begault, 1994; Wentzel, Wightman and Kistler, 1991), indicates that such systems are accurate enough for real-time applications.

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 it is critical to have a clear awareness of the geographic locations of every member. Maps (both static and dynamic) are one of the common 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 may cause him to miss critical visual information in the surroundings, or the physical environment does not permit visual navigation aids (due to smoke, darkness, mud, etc). Other people that may benefit from audio displays are for example the visually impaired.

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, Cheok, Yang and Qiu, 2004) has shown good results when navigating inside a room with a 3D audio cue. This study will however test if previous studies can be applied to a whole building. The hypothesis, that it is possible to successfully navigate and find objects using 3D audio information without the help of visual aids, was set in context by comparing user target searching by using 3D audio to either using a map or to do an exhaustive search.

METHOD

The experiment took place at the IT-University, Chalmers University of Technology, Gothenburg, on the second, third and fourth floors, see figure 1. Each floor covers about 40 x 40 meters. Three different conditions (independent variables) —no aid, 3D audio, and map— were compared with respect to the time the subject used for searching for a target (dependent variable).

Figure 1. Map over fourth floor as used by participants.

Participants

There were 24 participants in this study, 11 of which were female. The age span was 15 to 26 years. All participants were screened to have no hearing impairment, no physical impairment (for walking etc.) and no prior knowledge of the premises of the IT-University, Gothenburg.

Hardware and Software Development

To test the research hypothesis, a custom-made head tracker and software system were built and integrated with a commercial positioning software platform. The main parts of the head tracker were a microprocessor (BasicX-24, NetMedia), an electronic compass (HMR 3300, Honeywell) and an accelerometer (R2999M, Memsic). The head tracker had an update frequency of about 8Hz, and even though an update frequency of about 25 Hz is recommended, the overall performance of the system was found acceptable. Commercial Wireless Local Area Network (WLAN) positioning software from Ekahau Inc. was used to track the user and had an update frequency of about 0.3Hz and an accuracy of a few meters (often better). A pair of earphones was used for audio playback. See figure 2 for a picture of the hardware setup used in the user study.

Experimental Design

The study was designed as a between-subjects experiment with the conditions “no aid”, “map”, and “3D audio”. There were 8 subjects assigned to each condition. Each participant was tasked to search for hidden objects (see figure 3) one at a time. Each test session consisted of five such search tasks. In all conditions, the participant was told to find a specific, named object. In the “no aid” condition, the participant had no help finding the object. In the “map” condition, the participant had a map over three floors of the building and a mark where to find the object (see figure 1). In the “3D audio” condition, the participant was guided towards the object by a bell sound that was played in an endless loop sounding from the direction of the object. The 3D audio did not only provide the direction cue, but also a distance cue by changing the volume. If the user was on the right floor, there was a significant volume change and when the user was closing in on the target the volume was increasing. Also, the sound was played at +/-60 degrees elevation if the target was on a floor above/below. Each participant chose a HRTF using a procedure developed by Seeber and Fastl (2003). Participants chose from eleven HRTFs from the free CIPIC HRTF library (Algazi, Duda, Thompson and Avendano, 2001). The result of the procedure should result in eight out of ten subjects having the optimal or nearly optimal HRTF.

To avoid biases, all participants wore the same equipment and each participant practiced in a trial task before the actual experiment started. The order of the object targets was randomized for each participant to counterbalance systematic effects of practice. The completion time to find each object was measured and recorded.

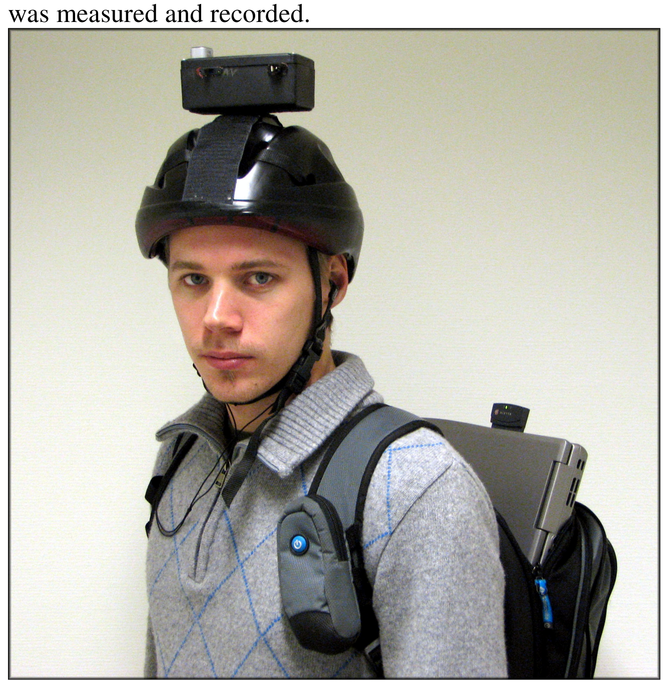

Figure 2. Hardware setup. Helmet with headtracker and backpack with laptop running all the software.

Procedure

First the participant was informed about the experiment and filled out a consent form. Next, the participant read a test scenario. In the 3D audio condition, the participant then chose an appropriate HRTF (Seeber and Fastl, 2003). The participant was then informed in which areas the objects could be found and that they should walk in a moderate pace during the tasks. After that the equipment were mounted on the participant. Then a trial task was conducted. During the trial task the participant was able to ask questions. When the object was found and the participant returned to the preparation room, the experimental phase begun. After each task the participant was distracted by filling in fraction papers while the assistant placed a new object in the building. Upon completion of the experiment, the participant was asked to fill out a questionnaire on their experience of the test and the aid used.

Figure 3. One of the objects to be found. Note the number in the bottom of the object - it corresponds to a number on the map aid.

RESULTS

To be able to draw statistically valid conclusions, an ANOVA and a Post-Hoc analysis was performed. The overall significance level was set at 5%. The variances were not equal between the groups.

The ANOVA analysis showed that the difference in time to find the objects for the conditions was significant ($F(2,117) = 30.61$, $p < .001$). Figure 4 shows a box plot of the average completion time for each condition. Using 3D audio, the average time to find the objects was 150 (s.d. 156.03) seconds. Using the map, the average time was 89 (s.d. 45.61) seconds. In the no aid condition, the average was 384 (s.d. 262.18) seconds.

Post-Hoc analysis (TAMHANE), 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 and the map was also 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%. No learning effects, sex or age differences were found in any of the conditions.

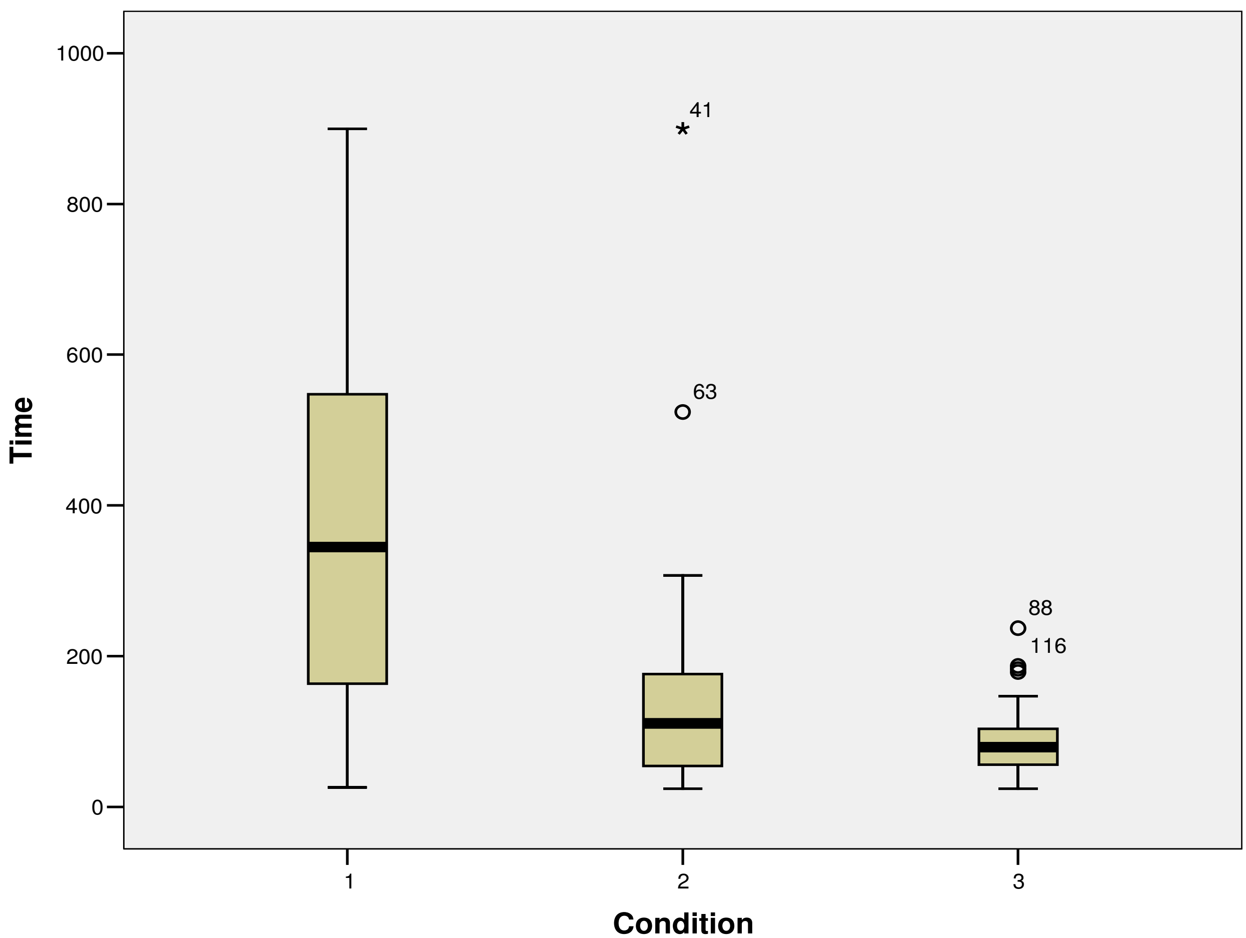

Figure 4. Box-and-whisker plot of average completion time (in seconds) for the search task.
Condition (1) is no aid, (2) is 3D audio and (3) is map.

The analysis of the questionnaire data does not differ between the map and the 3D audio other than that the map is rated very easy to use and the 3D audio is only rated fairly easy to use. The participants using the 3D audio aid rated the relative direction (i.e. the use of the head tracker) and the 3D perception as very important factors. The positioning accuracy was experienced as less important, and the system latency and the sound played was neither important nor unimportant.

The questionnaires and experimental assistant protocols indicate that many 3D audio users experienced that their mental focus on the audio is decreasing during the experiment and that they "learn to hear" over time. Such effects cannot be observed from the data set, however.

DISCUSSION

This study yields 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.

A main characteristic in the 3D audio condition is that the participants have trouble hearing elevation, but heard azimuth without problems. This resulted in participants searching on the wrong floor but at the "right" place on some occasions, as well as spending rather much time in the stairways listening for the volume cue. If this study was performed in a one-floor environment, performance using 3D audio beacons would most probably be better than for the map.

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.

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 and than the current position. Thus, a 3D audio presentation is very suitable to such a task. Other groups that might benefit from 3D audio include persons with visual impairments that can be guided by their ears rather than their eyes.

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.

FUTURE WORK

The results from this study recommend using different audio cues to represent different information in the audio domain. When navigating inside buildings, it would probably be a good idea to pitch the sound playback to indicate if the object is on another floor.

ACKNOWLEDGEMENTS

The authors would like to thank Ekahau Incorporated for the use of their WLAN positioning software and The Regents of the University of California for the HRTF library. Thank you Chalmers University of Technology and the IT-University in Gothenburg for giving us the possibility to present this paper at the HFES06 conference.

REFERENCES

ALGAZI, V. R., DUDA, R. O. THOMPSON D. M. AND AVENDANO, C. (2001). The CIPIC HRTF Database. *IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics 2001.*

BEGAULT, D. R., (1994). *3-D Sound for Virtual Reality and Multimedia*. Academic Press, London.

CHEN, F., CARLANDER, O. W. (2003). *Review on 3-Dimensional sound and head related transfer functions.* Linköping, Sweden: Swedish centre for human factors in aviation, Linköping institute of technology. ISBN: 91-7373-693-7.

KING, R. B., OLDFIELD, S. R. (1997). The impact of signal bandwidth on auditory localization: implications for the design of three-dimensional audio displays. *Human factors, 39(2)*, 287-295.

MURPHY, D. T., KELLY, M. C., AND TEW, A. I. (2002). 3D Audio in the 21 Century. In *Proceedings of the 8th International Conference, ICCHP*, Linz, Austria, 562-564.

SEEBER, B. U., FASTL, H. (2003). Subjective selection of non-individual head-related transfer functions. In *Proceedings of the 2003 International Conference on Auditory Display,* Boston, MA, USA, 1-4.

WALKER, B. N., LINDSAY, J. (2003). Effect of Beacon Sounds on Navigation Performance in a Virtual Reality Environment. In *Proceedings of the 2003 International Conference on Auditory Display,* Boston, MA, USA, 204-207.

WENZEL, E. M., WIGHTMAN, F. L., KISTLER, D. J. (1991). Localization with non-individualized virtual acoustic display cues. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology.* New Orleans, Louisiana, USA, 351 – 359.

ZHOU, Z., CHEOK, A. D., YANG, X., AND QIU, Y. (2004). An experimental study on the role of 3D sound in augmented reality environment. Interacting with Computers, 16(6), 1043-1068.