Spatial Sound Localization in an Augmented Reality Environment

Jaka Sodnik, Saso Tomazic
Faculty of Electrical Engineering
University of Ljubljana, Slovenia
jaka.sodnik@fe.uni-lj.si

Raphael Grasset, Andreas Duenser, Mark Billinghurst
Human Interface Technology Lab
University of Canterbury, New Zealand
raphael.grasset@hitlabnz.org

ABSTRACT
Augmented Reality (AR), the overlay of virtual images onto the real world, is an increasingly popular technique for developing new human-computer interfaces. As human navigation and orientation in different environments depend on both visual and auditory information, sound plays a very important role in AR applications. In this paper we explore users' capability to localize a spatial sound (registered with a virtual object) in an AR environment, under different spatial configurations of the virtual scene. The results not only confirm several previous findings on sound localization, but also point out some important new visual-audio cues which should be taken into consideration for effective localization and orientation in AR environment. Finally, this paper provides tentative guidelines for adding spatial sound to AR environments.

Author Keywords
Spatial sound, localization, augmented reality.

ACM Classification Keywords
H5.5. Information interfaces and presentation (e.g., HCI): Sound and Music Computing.

INTRODUCTION
Augmented reality (AR) involves the overlay of virtual imagery on the real world. It enhances the user's normal view of the world by adding computer-generated visual and auditory information. AR can be used in various application domains, such as visualization, medicine, engineering and education (Azuma, 1997; Azuma, 2001).

Most previous research in AR environments has been concerned with overlaying virtual graphics on the real world. In contrast, we are interested in audio enhancements and how spatial sound can combine with graphics to improve performance in an AR application.

This combining of 3D virtual graphics and spatial sound in AR not only offers a new type of applications but also new possibilities for the audio research community. For example, a user could move a tangible object (e.g. cardboard cube) on a table and hear a 3D sound of a motor engine while seeing a virtual model of a car moving with it. Natural interaction with real objects and the environment, and seamless merging between real and virtual content, introduce new research problems which differ from immersive virtual reality. For sound researchers the possibility to freely register and spatially configure the position of sounds with virtual images has appeal for perceptual studies.

In this paper, we conduct an evaluation for AR visual and sound localization in the context of tabletop situations. This is an area that has not been well studied in the past, and this paper provides guidelines that will be useful for AR interface developers.

RELATED WORK
Spatial sound has been proven to play an important role in AR applications. Its use has been explored in very different areas, from pure entertainment (Stampfl, 2003a) to video conferencing and remote collaboration (Billinghurst 2001, Regenbrecht, 2004), and also perceptual studies (Bormann , 2005). In this section we review earlier work in adding spatial sound to AR environments.

AR Spatial Sound Environment
The majority of previous research that studied the use and importance of spatial sound in user interactions focused manly on Virtual Reality (VR) and desktop computer environments.

Billinghurst used VR techniques for information representation using spatial sound in a wearable computer interface (Billinghurst, 1998). He established that body-stabilised displays provide benefits over traditional head-stabilised displays, especially when enhanced with spatial audio and visual cues. In this case, a simple spatial sound cue enabled users to find specific pieces of information in a visual search task more effectively.

Teleconferencing and remote collaboration using VR is a potential application area for 3D sound. Regenbrecht's cAR/PE videoconferencing system comprises of live video streams of the participants arranged around a virtual table with spatial sound support (Regenbrecht, 2004). Spatial sound driven by headphones or 2.0 to 7.1 audio hardware was used to indicate different user positions. Usability studies showed general usability as well as good overall satisfaction of the users.

OzCHI'06, November 22-24, 2006, Sydney, Australia.

Parameterization of sounds using timbre trees in virtual environments was explored by Hahn (Hahn, 1998). He presented an integrated system for modelling, synchronizing and rendering sounds for virtual environments. In this system, the sound parameters are mapped to the parameters associated with the motions of objects in environment. The goal of the research was to observe correspondence of motions and sounds in the virtual environment.

Based on these experiments, other researchers have also explored how spatial sound can be used in Augmented Reality environments.

Haller, Dobler and Stampfel built an interesting low-cost AR interface for positioning musical instruments in space in real time (Dobler, 2002; Haller, 2002). The users can move and manipulate sound sources around themselves in space with a special pen and observe the results immediately. Sound sources are represented with 3D models of instruments and spatialized audio appears to be playing from the location of the virtual instruments.

AR/DJ is another example of an AR interface for manipulating 3D sound (Stampfl, 2003b). It is an application that allows two music DJs in a club to play many different sound samples and place them anywhere in 3D space in the club. Sound sources are visualized in a 3D model of the dance floor and they can be manipulated using a pen with the visual tracking markers on it.

Stampfl (Stampfl, 2003a) also developed the 3deSoundBox which is a platform-independent acoustic component for driving any number of speakers, which works with various applications on various platforms and has a very scalable architecture. Its main task is to provide a platform for exploring new possibilities in the field of virtual and augmented reality applications.

Another interesting example of the use of sound in an AR interface is an interactive audio museum guide (Hatala, 2004). In this case, the visitor's location within the museum is tracked and dynamic audio data played back related to the artefacts the visitor is seeing. Using gestures the visitor can interact with a single artefact or multiple artefacts (3D audio) in order to listen to related audio information. The interface enables users to interact with the system by movement and object manipulation-based gestures.

Localization Experiments with AR and Spatial Sound

Several researches reported using Augmented Reality for sound localization experiments.

Zahorik described a study of the role of visual-feedback training in 3D sound localization (Zahorik, 2002). In the study he wanted to find out if perceptual training can reduce localization errors caused by the use of low-cost 3D audio equipment and non-individualized head related transfer functions (HRTFs) (Wang, 2002). Paired auditory/visual feedback was provided to the listeners through a head mounted display (HMD).

Sundareswaran described a 3D audio wearable system which could be used to provide alerts and informational cues to a mobile user (Sundareswaran , 2003). Mobile users were able to navigate in a virtual environment based on spatial sound cues. An experiment was conducted to observe the role of feedback training on sound localization accuracy.

Bormann tried to establish if high fidelity audio leads to higher feelings of user presence (Bormann, 2005). He varied the fidelity of spatial sound while performing search tasks in a virtual reality environment. He reported that lower audio fidelity resulted in both the lowest performance and the highest increase in user presence.

The most relevant work is the study on the impact of 3D sound on depth perception in an augmented reality environment (Zhou, 2004). Zhou reported significant improvement of depth perception of virtual objects, when spatial sound was also present.

Our Research Contribution

The majority of previous research was focused on a large room scale environment, with little work studying 3D sound in a tabletop AR environment. Tabletop AR applications are an important class of AR interfaces for collaborative work, gaming, architecture or engineering developments.

The aim of our research is to evaluate the perception and localization of 3D sound in a tabletop AR environment. A major problem with this type of application is poor virtual object distance or depth perception, especially if the objects are located in near proximity (Zhou, 2004). The problems we are therefore focusing on are:

1/ Will the localization of a sound combined with a visual object have an impact on the users' perception?

2/ Will the short distance perception in an AR tabletop configuration give different results from previous studies?

Our work is different from previous AR spatial sound research in a number of ways:

- it is conducted on a tabletop environment
- the evaluation of localization performance is based on visual and aural cues
- individual localization cues are isolated

In the rest of the paper we describe our user study. We first give an overview in the User Study section, then the practical realization and technical background is given in the Methods section. The results of the experiment are interpreted in detail and the impact of different localization cues is pointed out. The Discussion section summarizes the important findings, and we end the paper with ideas for future research.

USER STUDY

We are interested in combining visual and sound cues for navigation and localization in tabletop AR environments. In our user study all possible locations of sound sources are marked with identical 3D virtual models. Spatial sound could be attached to any of the models at certain times. Since all models are identical, the visual cues can only serve for micro-orientation while the sound cues aid macro-orientation. In other words, users should first

navigate according to the sound and when the approximate location of the sound source is found, the position of the 3D model should help to distinguish the exact location of the source. The goal of the research is to explore sound-visual navigation in a number of conditions, whereby different localization cues can be compared and evaluated.

This experiment will help inform interface designers how spatial sound cues combined with visual cues can be used to improve localization in AR environments.

Apparatus

An AR scene approximately 100cm x 60cm x 60cm was created which was observed through a HMD with an attached video camera. The scene consisted of 24 identical models of a small Cessna airplane. All 24 models could be seen at the same time and were spatially configured in four rows and six columns (six airplanes in each row) (see figure 1). Beneath the models a simple two-coloured virtual ground plane was shown.

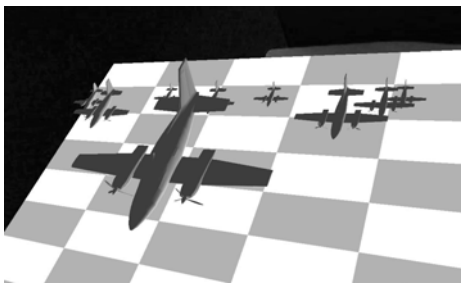

Figure 1. Virtual scene with airplanes

Using the ARToolKit computer vision tracking library (Kato, 1999), the virtual scene was overlaid on a real piece of paper with tracking markers drawn on it. Users could see the virtual object cues at the same time as the real world. The paper was placed on a table and the user was seated behind the table on a rotating chair (see figure 2).

Figure 2. User performing the localization test

The user could move the chair in any direction, and rotate their head or body in order to move around in the virtual scene. In order to hear the spatial sound cues, the user also wore stereo headphones.

Task

In the experiment the user was asked to find which of the airplanes was generating a sound cue. He or she was supposed to lean in and approach the target airplane as close as possible to validate this supposition. The selection tolerance between the user's head and the plane was 15 cm, so when the user was within 15cm of the sound source the task was completed. The sound stimulus was a sequence of white noise (described in "Stimulus"), simulating the motor engine sound.

Conditions

There were 5 different conditions, based on 3 different spatial configurations that we describe below.

Configuration 1: Horizontal

In this case all airplanes were located on a horizontal plane (see Figure 3). That means that there was no difference in elevation between the models. The distance between the models and the user's head varied from approx. 15cm to 80cm. There were 24 airplanes in total laid out in four rows of six airplanes.

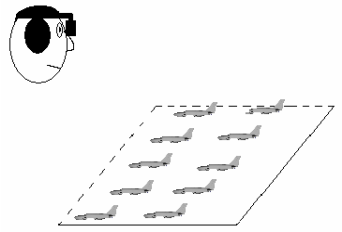

Figure 3. Configuration 1 – horizontal plain

Configuration 2: Vertical

In the second configuration, each row of airplanes was located at different elevations (see Figure 4). There was 15cm of vertical distance between each row of airplanes. The elevation was increasing towards the user which means that the row of airplanes closest to the user was located at the highest elevation. As in the previous condition there were four rows of six airplanes.

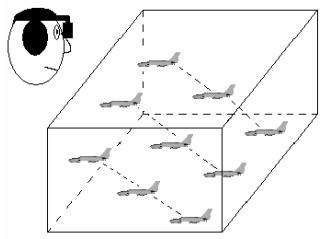

Figure 4. Configuration 2 – rising elevation

Configuration 3: Random

In the last configuration, the airplanes were still in four rows with six airplanes, but this time the elevation of individual airplane was random (see Figure 5). The airplane height varied between 15cm and 60 cm above the virtual ground plane.

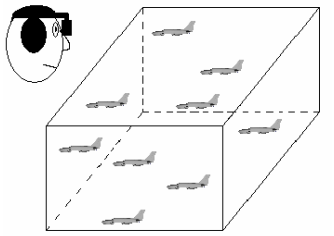

Figure 5. Configuration 3 – shuffled elevation

Using these three configurations we created five experimental conditions:

Condition 1; Horizontal - HRZT: The airplanes were arranged in configuration 1. The user was asked to try to find only the correct column. That means that only horizontal or azimuth localization was preformed. This has previously been proven to be very efficient with the use of non-individualized HRTFs (Wenzel, 1993; Sodnik, 2005).

Condition 2; No Elevation – NOEL: Airplanes were still arranged in configuration 1, but this time the user was asked to find only the "noisy" airplane.

Condition 3; Rising elevation – RIEL: The airplanes were arranged in configuration 2 and the user was asked to find the "noisy" airplane.

Condition 4; Shuffled elevation - SHEL: The airplanes were arranged in configuration 3 and the user was asked to find the "noisy" airplane.

Condition 5; Coded elevation - COEL: This condition was almost identical to the fourth condition, but in this case "artificial elevation coding" was used in order to improve vertical resolution. The details and the background on artificial elevation coding are explained next in the Methods section.

METHODS

Subjects

Ten unpaid volunteers participated in our experiment (six men and four women with ages ranging from 21 to 35 years, and a median age of 27.6 years). They all reported normal eyesight and hearing, and none of them had participated in a sound localisation experiment before.

Procedure

A within-subjects design was used. Each user, after being positioned and equipped, was instructed to try to find the "noisy" airplane. They were given five warm-up localization trials in condition 1. No hints or instruction were given to the users on how to perform the task.

In each condition the users were supposed to localize the model with the sound cue in less than five attempts. When they thought they were close to the sound source, they clicked a mouse button to confirm the selection. If target localization was successful (they were less than 15 cm away from the correct target), a new target was selected (the new "noisy" airplane was always selected randomly). After five unsuccessful attempts a new target was selected anyway. In each condition, the user had to perform ten localization tasks.

The number of attempts was recorded each time. Every time the mouse button was pressed, the current distance between the target and the user was recorded in a log file. To reduce order effects, each condition was experienced in a counterbalanced manner.

At the end of the experiment, the users were asked to rank the five conditions on a scale from 1 to 5, according to how difficult the task was (5 being the easiest and 1 being the hardest). We also collected some observations of the users' behaviour while preforming the tasks.

The experimental measures collected were the following:

- The average number of attempts to find the sound source
- The distance between the user and the targeting object at each attempt
- Subjective evaluation of the conditions by the users

Our expectation is that localization performance should differ significantly in different conditions. In the first three conditions we need to observe the capability of azimuth, distance and elevation perception respectively. Azimuth perception of sound sources can be quite accurate with non-individualized HRTFs (Wenzel, 1993), so we expect that in an AR scene it should be significantly better than perception of the other two directions. On the other hand, visual cues should also enable correct perception of distance and elevation. The fourth and the fifth condition are expected to be the most difficult tasks, since objects are randomly distributed in space. The simultaneous presence of visual and sound cues should allow the users a certain degree of learning and improve the localization performance.

A within subject analysis of variance (ANOVA) test establishes significant differences in the results of different conditions. Individual conditions are compared with post-hoc Bonferroni tests.

Design

The main parts of the AR application are the visual augmented reality environment (tracking and display) and the 3D sound reproduction. A calibration step is also necessary to align the visual and aural elements. In this section we describe these components in more detail.

Environment

The AR environment was based on a video see-through visualization using the ARToolKit computer vision tracking for the registration. A set of 41 black and white markers in the same plane (each 5cm x 5cm) was used for the user viewpoint position and orientation tracking. Each marker was approximately 10cm apart. During the experiment they were not visible to the user since they were covered with the virtual ground plane.

For viewing the AR scene an eMagin Z800 3DVISOR head mounted display was used, equipped with a Logitech QuickCam for Notebooks pro camera. The eMagin HMD was connected to a PC that had an NVIDIA GeForce 6800 GT graphic card.

The graphics application was written in C++, using the OpenSceneGraph (OpenSceneGraph) rendering library, and the OSGART (high level AR framework) for marker tracking and camera calibration. The application was developed under Windows and ran on a standard PC.

Sound Reproduction

For sound reproduction the Creative Sound Blaster X-Fi ExtremeMusic sound card with AKG K-44 headphones was used. Spatial sound generation was driven by the Creative OpenAL sound library (OpenAL) which enabled access to all X-Fi hardware accelerated 3D sound features.

OpenAL enables the simple positioning of virtual sound sources in 3D space using CMSS-3D surround sound technology on the Creative sound card. CMSS-3D contains the non-individualized Head Related Transfer Function (HRTF) library with direct support for playing through headphones.

Artificial Coding of Elevation

We previously mentioned the notion of "artificial coding of elevation". Originally, using CMSS-3D technology for 3D sound positioning, we depend on the accuracy of non-individualized HRTFs. Using the OpenAL API, virtual objects at different elevations are filtered with these HRTFs and the perception of elevation is therefore quite poor. The idea of artificial coding is to add some spectral cues to the signal to improve elevation perception. Based on reports of researchers who studied elevation localization in detail (Algazi, 2001; Rogers, 1992; Susnik, 2005), a simple low-pass filtering was applied on top of signals filtered with HRTFs. The cut-off frequency f_{cutoff} of the low-pass filter was changed according to the current elevation (the position of the listener's head relative to the selected object):

$$f_{cutoff} = f_{cutoff\_\min} + f_{cutoff\_\max}\left(1 - \frac{el_x - el_{\min}}{el_{\max} - el_{\min}}\right)$$

The variables el_{max}, el_{min} and el_x are the maximum, the minimum and the current elevation of the listener's head according to the selected virtual object. The values f_{cutoff_min} and f_{cutoff_max} are the maximum (20.000 Hz) and the minimum (2.000 Hz) cutoff frequencies.

As a result, virtual objects at low elevations sounded "low" since only low frequencies were contained in the spectrum. On the other hand, the higher the object was located (closer to the user) in the scene, the wider the frequency spectrum.

Stimulus

In all experiments, a repeating sequence of white noise (one second long) served as the stimulus. White noise was chosen because it has been shown to be the most suitable stimulus for sound localization due to its flat frequency spectrum (Susnik, 2003). White noise is also very easy to manipulate with a low-pass filter.

Calibration

Using ARToolKit, information about the current position of the camera (user's viewpoint) could be acquired at any time. The distance between the individual virtual object and the user can be calculated according to the relative position of the objects within the scene.

Since the dimensions of our virtual space were quite small, there was originally only a minor difference in the sound volume between two airplanes in the neighbouring rows. In order to improve distance perception, the difference in sound volume was exaggerated so that maximum volume was reached at approximately 15cm.

The current sound volume v_x was set manually as:

$$v_x = v_{\max}\left(1 - \frac{d_x - d_{\min}}{d_{\max} - d_{\min}}\right)$$

Here v_{max} is the maximum volume (1.0) of the sound source, while d_x, d_{max} and d_{min} are the current, maximum and the minimum distances to the sound source.

RESULTS AND INTERPRETATION

Number of trials

Figure 6 shows the average numbers of attempts to find the sound source for each of the five localization conditions.

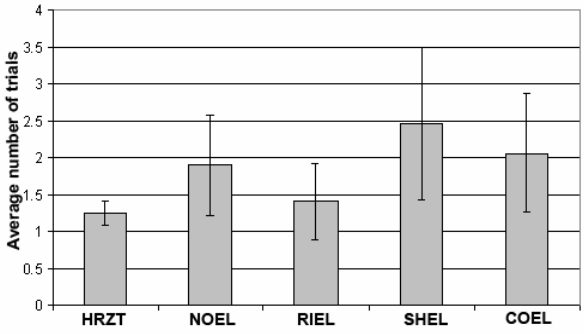

Figure 6. Average number of trials in different conditions

There was significant difference between the results of individual conditions. A within subject ANOVA test, resulted as: $F(4,36)=6.759$, $p < 0.001$.

The first HRZT condition has the lowest average number of attempts. The confidence interval is also quite narrow, which signifies a high localization accuracy. This is in accordance with the fact that azimuth localization is very accurate although non-individualized HRTFs are used (Sodnik, 2005). A post-hoc comparison (Bonferroni) between HRZT and the other four conditions found a significant difference between the mean values of HRZT and SHEL ($p < 0.01$) and HRZT and COEL ($p < 0.01$).

A post-hoc (Bonferroni) comparison between other conditions did not find any significant difference in the mean values.

The distance to the target

During the localization tests the distance of the user's viewpoint to the target when the mouse was clicked was

also recorded. Using this we can observe how the users' distance to the target decreases within five attempts. A distance of 150mm was the limit below which the targeting objects was considered to be localized. Figure 7 shows the average distances to the targets for the first three attempts. The distance is calculated as an average of 10 trials.

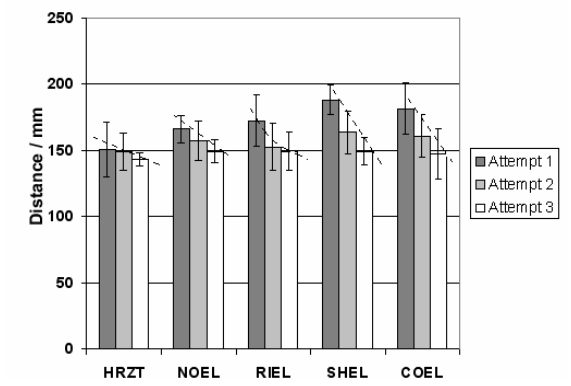

Figure 7. The distances to the target at first three attempts

The difference in the heights of the bars in each condition shows the learning effect of the test population within each trial. The dashed line at the top of the bars thus represents the learning curve. The within subject ANOVA showed significant difference between individual conditions: $F(4,36)=33.06$, $p < 0.01$.

In the HRZT condition there is not much improvement between the attempts because the distance is already very small at the beginning. The gentle learning curve shows a weak learning effect. Also ANOVA test confirmed non-significant difference between the attempts: $F(2,18)=1.692$, $p = 0.21$.

The starting distance in the NOEL condition is the second lowest and also does not decrease very steeply. That means that the users were able to quickly approach the targeting source but then they had problems finding the exact model. The ANOVA reported significant difference between the attempts: $F(2,18) = 4.532$, $p = 0.25$. The post-hoc Bonferroni comparison reported significant difference between the first and third attempt ($p = 0.02$).

The RIEL condition is very interesting, since it has a relatively high initial value and high gradient at the beginning. There is a significant decrease of distance to the target between the first and the second attempt, but only a slight decrease to the third one (ANOVA test for all three attempts: $F(2,18) = 14.939$, $p < 0.01$). That means that the users were quickly able to find the way to localize the correct model. The post-hoc Bonferroni test showed that the distance at first attempt differs significantly from the other two attempts ($p < 0.01$; $p < 0.01$).

The SHEL and COEL conditions are quite similar with learning curves with a high gradient. In both cases there was a big error or long distance to the target at the beginning followed by a fast learning effect. With artificial coding a better initial position was achieved and the correct model was localized quickly. In both cases ANOVA confirmed significant difference between the trials: SHEL: $F(2,18) = 63.558$, $p < 0.01$; COEL: $F(2,18) = 10.770$, $p < 0.01$. The post-hoc Bonferroni test also confirmed significant difference between the three trials in SHEL condition (1. and 2.: $p < 0.01$; 2. and 3.: $p = 0.33$; 1. and 3.: $p < 0.01$). In the COEL condition only 1st attempt is significantly different from the other two (1. and 2.: $p = 0.02$; 2. and 3.: $p = 0.44$, 1. and 3.: $p = 0.25$).

We also studied if there was a learning effect along the 10 trials for each condition. We restricted this analysis to the observation of the average distance to the target in the first attempt. Figure 8 shows the results of all users in the five conditions.

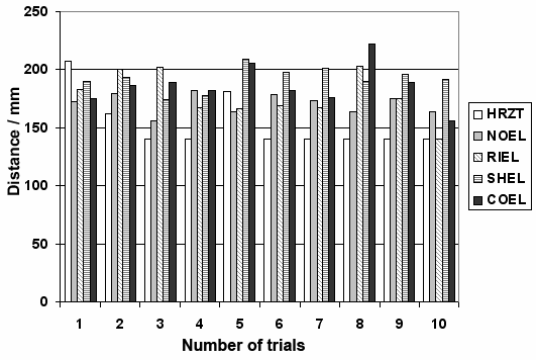

Figure 8. The distances to the target at the first attempt during 10 trials

It can be seen that there is no learning effect during the 10 localization trials. The starting distance to the target in the users' first attempt is not decreasing but is more or less random.

Subjective ranking of the five conditions

At the end of the experiment the users were asked to evaluate the difficulty of individual conditions. They ranked all five conditions on a scale from 1 to 5, where 5 was the easiest condition and 1 was the hardest condition. Table 1 shows the average rank of the different conditions.

HRZT	NOEL	RIEL	SHEL	COEL
4.6	2.4	4.3	1.3	2.4

Table 1. Average ranks of the five conditions

From the results in table 1 we can conclude that, in general, the accuracy of localization correlates with its subjectively established difficulty. It seems that the users ranked the difficulty of different conditions according to their success in the trials. The only exceptions are the NOEL and COEL conditions that were given the same rank by the users although the localization accuracy of the latter is somewhat lower. The non-parametric Friedman test showed significant difference between individual conditions: $\chi^2(4, N = 10) = 31.44$, $p < 0.01$. A post-hoc Bonferroni confirmed that HRZT and RIEL conditions do not differ significantly ($p < 1.000$), but they are significantly different from NOEL ($p < 0.01$; $p < 0.01$), SHEL ($p < 0.01$; $p < 0.01$) and COEL ($p < 0.01$; $p < 0.01$) conditions.

In the NOEL condition, all models and sound sources were in the same plane. That means that beside the azimuth localization only the sound volume could help with the perception of distance to the object. Even with moving their head and body, the users were sometimes unable to localize the targeting model. In the "rising elevation" case (RIEL), the models were located at different heights but still quite systematically arranged. Difference in height or elevation served as an additional cue for sound localization. The users were able to move their head around the models and through them and so localize the sound easily.

The difference between RIEL and SHEL can be interpreted as the impact of the a systematic arrangement of elevation, which helped the users in performing their task. When the models were positioned in space randomly, it was harder for the users to find some systematic way to search through them. Condition SHEL proved to be the most difficult for all users.

NOEL, SHEL and COEL were not significantly different from one another (NOEL-SHEL: $p < 0.243$; NOEL-COEL: $p < 1.000$; SHEL-COEL: $p < 0.174$).

With the COEL condition, our attempt was to improve the elevation cue with artificial sound coding. Most of the users confirmed that unnatural behaviour of the sounds at different elevations was very confusing at the beginning, but after few trials they reported that it was helpful.

The observation of the users while they were performing the localization tasks showed that they mostly tried to localize the target in the horizontal plane first. The users achieved this by moving their head or body left and right until the sound appeared to be coming from directly in front of them. After establishing the azimuth of the sound, the users proceeded to localize the individual model in the column.

DISCUSSION

In our pilot study with two users we observed the time necessary for accomplishing each localization task. There were big differences in localization time between users, between different conditions and also between individual trials in each condition. Because of the randomness of time variations, we decided to exclude the observation of time from the main experiment.

The results of the first condition confirmed that azimuth of arbitrary sound source can be located with high accuracy, although non-individualized HRTF filters are used (Wenzel, 1993, Sodnik, 2005).

The number of subjects limited us from performing a deeper quantitative analysis of our experiments. This study should be repeated with more test subjects to confirm our pilot results. However, in this paper we describe the first interpretation of these elements in the following paragraph.

The second condition was the evaluation of depth or distance perception of spatial sound, since all models were located in the horizontal plane. The analysis of the results confirmed that the distance perception of near sources is poor (Zhou, 2004). When sound sources are in close proximity, elevation can play an important role for better perception. The latter was confirmed especially with the third condition of our experiment, whereby the models were equally distributed in the vertical direction. This condition proved to be the most ideal of all for localization and was ranked as the easiest by the users. When the vertical distribution was random, the localization performance dropped. The users reported this task to be the hardest one. With the last condition we attempted to show that localization of randomly distributed objects can be improved with artificial coding or spectrum manipulation. At this stage, we did not get any significant results to demonstrate the claim. Based on good results of previous studies on artificial coding of elevation in the acoustic image (Susnik, 2005), we expect that also in AR environment the elevation perception could be improved in this way.

The measurement of distances to the targets enabled the observation of learning effects in two different dimensions. There was a noticeable learning effect within each of the ten trials. Comparing the distances in the first, second and third attempts (figure 7), we can observe the users' technique in localizing individual sources. The user's first attempt was based specifically on sound cues when the user macro-localized the target. The micro-localization (the second and third attempt) was based on visual cues. There was a significant difference in learning effect between conditions.

On the other hand, there was no learning effect while performing individual trials in each condition. The distance to the target on the first attempt was related to the sequence of the trial (figure 7). Sometimes the starting distance in the later trials (7, 8, 9) was much greater than in the first few trials (1, 2, 3).

Design recommendations

Based on the results of the experiment, we can give some suggestions for the design and development of tabletop AR applications with spatial sound. Non-individualized HRTF libraries enable satisfactory sound localization in AR environments when visual cues are also present. Very accurate azimuth localization of the sound can be achieved, and poor elevation and distance perception can be enhanced with the addition of visual cues. That means that off-the-shelf sound cards with in-build HRTF filters can be used for effective sound perception in AR environments. Their major advantage is the simplicity of use, since they can be driven with simple positional libraries (OpenAL) and therefore used in entertainment or gaming. In order to achieve accurate spatial sound perception in all dimensions, an artificial elevation coding technique might be used. Beside the low-pass filter technique also other artificial coding techniques can be applied (oscillators with different central frequencies, pitch changing, etc.)(Susnik, 2005).

CONCLUSION

Spatial sound represents an important cue for navigation in space. This experiment explores the possibility of localization in a tabletop augmented reality environment,

based on sound and visual cues. We used five different configurations of 3D models to evaluate the impact of different components of spatial sound on the localization performance.

The results of our experiment confirm that humans localize the azimuth of sound source much better than elevation or distance. Localization performance is especially poor when targeting objects are randomly distributed in space. Our experiment shows that localization can be improved if there is some regularity in the distribution. The improvement of elevation perception can also be made with artificial coding. This proved to be an effective method, but it requires some learning,, and it could also sometimes appear to be unnatural and disturbing. We are planning to perform further research into this problem.

This evaluation shows the results of localization of virtual sounds combined with virtual objects. In the future we plan to compare the importance of contradictory virtual and acoustic cues for navigation, i.e. the virtual images and audio cues which do not correspond to one another. We are going to try to confuse the users performing the localization task by not attaching the sounds to 3D models. We are also interested in further exploring the usage of 3D sound for AR applications and how to use perceptual factors to improve the sound rendering.

REFERENCES

Algazi, V.R., Avendano, C., Duda, R.O. Elevation localization and head-related transfer functions analysis at low frequencies. Journal of the Acoustical Society of America 109, 3 (2001), 1110-1122.

Azuma, R.T. A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6, 4 (1997), 355-385.

Azuma, R., Baillot Y., Behringer R., Feiner S., Julier S., MacIntyre B. Recent Advances in Augmented Reality. IEEE Computer Graphics and Applications 21, 6 (2001), 34-47.

Billinghurst, M., Bowskill, J., Dyer, N., Morphett, J. Evaluation of Spatial Interfaces for Wearable Computers. In Proc. VRAIS '98, (1998).

Billinghurst, M., Kato, H., Poupyrev, I. The MagicBookc – moving seamlessly between reality and virtuality. IEEE Computer Graphics and Applications 21, 3 (2001), 6-8.

Bormann, K. Presnece and the Utility of Audio Spatialization. Presence 14, 3 (2005), 278-297.

Dobler, D., Haller, M., Stampfl, P. ASR – Augmented Sound Reality. In Proc. SIGGRAPH 2002, ACM Press (2002), 148.

Hahn, J. K., Fouad, H., Gritz, L. and Lee, J.W. Integrating Sounds and Motions in Virtual Environments. Presence 7, 1 (1998) 67-77.

Haller, M., Dobler, D., Stampfl, P. Augmenting the Reality with 3D Sound Sources. In Proc. SIGGRAPH 2002, ACM Press (2002), 65.

Hatala, M., Kalantari, L., Wakkary, R., Newby, K. Information access and retrieval (IAR): Ontology and rule based retrieval of sound objects in augmented audio reality system for museum visitors. In Proc. 2004 ACM Symposium on Applied Computing, ACM Press (2004), 1045-1050.

Kato, H., Billinghurst, M. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In Proc. IWAR 99, (1999).

OpenAL, http://www.openal.org/

OpenSceneGraph, http://www.openscenegraph.org/

Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M., Wilke, W., Mueller, E. Using Augmented Virtuality for Remote Collaboration. Presnece 13, 3 (2004), 338-354.

Rogers, M.E., Butler, R.A. The linkage between stimulus frequency and covert peak areas as it relates to monaural localization. Perception and Psychophysics 52, (1992), 536-546.

Sodnik, J., Susnik, R., Stular, M. and Tomazic S. Spatial sound resolution of an interpolated HRIR library. Elsevier, Applied Acoustics 66, 11 (2005), 1219-1234.

Stampfl, P. 3deSoundBox – a Scalable, Platform-Independent 3D Sound System for Virtual and Augmented Reality Applications. In Proc. EUROGRAPHICS 2003, (2003a).

Stampfl, P. Street tech: Augmented reality disk jockey: AR/DJ. In Proc. SIGGRAPH 2003, ACM Press (2003b).

Sundareswaran, V., Wang, K., Chen, S., Behringer, R., McGee, J., Tam, C., Zahorik, P. 3D audio augmented reality: implementation and experiments. In Proc. ISMAR '03, ACM Press (2003), 296-297.

Susnik, R., Sodnik, J. and Tomazic, S. Sound source choice in HRTF acoustic imaging. In HCI International adjunct proceedings, (2003), 101-102.

Susnik, R., Sodnik, J. and Tomazic, S. Coding of elevation in acoustic image of space. In Proc. ACOUSTICS 2005, (2005), 145-150.

Wang, K., Sundareswaran, V., Tam, C., Bangayan, P. and Zahorik, P. Efficient and Effective Use od Low-Cost 3D Audio Systems. In Proc. ICAD (ATR) 2002.

Wenzel, E.M., Arruda, M. and Kistler, D.J., Wightman, F.L. Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America 94, 1 (1993), 111-123.

Zahorik, P., Tam, C., Wang, K., Bangayan, P., Sundareswaran, V. Localization accuracy in 3-D sound displays: The role of visual-feedback training. In Proc. ICAD (ATR) 2002.

Zhou, Z., Cheok, A.D., Yang, X., Qiu Y. An experimental study on the role of 3D sound in augmented reality environment. Elesvier, Interacting with Computers 16, (2004) 1043-106.