Available online at www.sciencedirect.com
SCIENCE DIRECT®
Interacting with Computers 16 (2004) 989–1016
Interacting with Computers
www.elsevier.com/locate/intcom

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

An experimental study on the role of software synthesized 3D sound in augmented reality environments

Zhiying Zhou, Adrian David Cheok*, Xubo Yang, Yan Qiu

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576

Received 19 January 2004; revised 17 April 2004; accepted 17 June 2004
Available online 5 August 2004

Abstract

Investigation of augmented reality (AR) environments has become a popular research topic for engineers, computer and cognitive scientists. Although application oriented studies focused on audio AR environments have been published, little work has been done to vigorously study and evaluate the important research questions of the effectiveness of 3D sound in the AR context, and to what extent the addition of 3D sound would contribute to the AR experience.

Thus, we have developed two AR environments and performed vigorous experiments with human subjects to study the effects of 3D sound in the AR context. The study concerns two scenarios. In the first scenario, one participant must use vision only and vision with 3D sound to judge the relative depth of augmented virtual objects. In the second scenario, two participants must co-operate to perform a joint task in a game-based AR environment.

Hence, the goals of this study are (1) to access the impact of 3D sound on depth perception in a single-camera AR environment, (2) to study the impact of 3D sound on task performance and the feeling of 'human presence and collaboration', (3) to better understand the role of 3D sound in human-computer and human–human interactions, (4) to investigate if gender can affect the impact of 3D sound in AR environments. The outcomes of this research can have a useful impact on the development of audio AR systems which provide more immersive, realistic and entertaining experiences by introducing 3D sound. Our results suggest that 3D sound in AR environment significantly improves the accuracy of depth judgment and improves task performance. Our results

* Corresponding author. Tel.: +6874-6850; fax: +6779-1103.
E-mail address: adriancheok@nus.edu.sg (A.D. Cheok).

0953-5438/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.intcom.2004.06.014

Meta Platforms, Inc.
Exhibit 1107
Page 001

also suggest that 3D sound contributes significantly to the feeling of 'human presence and collaboration' and helps the subjects to 'identify spatial objects'.
© 2004 Published by Elsevier B.V.

Keywords: 3D sound; Augmented reality; Depth perception; Localization of sound; Task performance; Collaboration; User studies

1. Introduction

Audio feedback, especially 3D sound, has proven itself to be an effective addition to human–computer interfaces in many studies. Sounds can be used to present information that would otherwise be difficult to convey through other means (Gaver et al. 1991). Moreover, 3D sound can convey location information for specific alerts or warnings. Also, 3D sound was reported to increase the immersive experience in virtual reality (VR) environments and hence provide a more realistic virtual environment (Begault 1994). Furthermore, the addition of 3D sound has been shown to increase the performance in active tasks (Wenzel, 1996; Avionique). In previous works, Mereu (1995) suggested that sound could improve depth perception in 3D interfaces. Makino et al. (1996) also indicated that spatialized sound from virtual acoustic display helps fast and accurate navigation. Additionally, simulations by Gaver et al. (1991) suggested that auditory icons can play a significant role in multiprocessing and collaborative system. Unlike VR in which extensive research has been carried on 3D sound, augmented reality (AR) aided by 3D sound is still a relative new research area.

AR, as placed on the reality–virtuality continuum (Milgram et al., 1994), refers to the incorporation of virtual computer graphics objects into a real world scene, or alternatively the inclusion of real world elements into a virtual environment. The former case is generally referred to as AR, and the latter as augmented virtuality. Azuma (1997) has defined three characteristics that are integral to an AR interface. First, it combines the real and the virtual. Second, it is interactive in real time. Third, it is registered in three dimensions. Although AR is often associated with visualization (starting with the first head-mounted display by Sutherland (1968)), augmentation is also possible in the aural domain (Cohen et al., 1993; Mynatt et al., 1997). If we extend this notion to include sonic effects, a real-time registered three dimensional (3D) sound could also be a paradigm of AR applications, and is termed augmented audio reality (AAR) (Cohen et al., 1993).

A few applications have applied 3D sounds in AR to enhance user experiences. Billinghurst et al. (1998) presented a prototype wearable communication space that uses spatial visual and audio cues to enhance communication between remote groups of people. The 'Listen Project' (Eckel et al., 2001) provides users with intuitive access to personalize and situated audio information spaces while they naturally explore everyday environments. It augments the physical environment through a dynamic sound-scape, which users experience over motion-tracked wireless headphones. The 'Magic Music Desk' (Zhou et al., 2002) registered virtual sound sources on fiducial markers from which 3D position information was retrieved by vision-based tracking method (Kato and Billinghurst, 1999). 'Touch space' (Cheok et al., 2004) applied different 3D sounds to represent different

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 002

moving avatars in AR environment. Although there are quite a few AAR applications, the effectiveness of 3D sound in AAR has been rarely investigated. The particular research problems and questions that arise are:

1. Exact spatial 3D sound in virtual and augmented environments requires perceptual externalization of the sound source, which can only be brought about by using head related transfer functions (HRTFs) (Bill G, 1994) (these are tuned to each individual), reflections, reverberation, etc. (Wenzel, 1994) which might be too complex to be included in normal real time AR applications (which need to be a real time interactive system for any user without individual training). An alternative way to generate 3D sound is to use open source software development kits (SDKs) which have already incorporated the general HRTFs database and simple model of environment to generate 3D sound. The minimum requirement of these SDKs to generate 3D sound is to update the relative position of the virtual sound source to the head co-ordinate system. Thus, with such sound sources, a problem turns to be how can we track the sound source relative to user's head co-ordinate system.
2. How can we precisely register virtual sound sources in the physical world? More specifically, what kind of tracking method is suitable for obtaining precise position and orientation of the sound source in AR environments?
3. As 3D sound has been reported to improve depth perception (Mereu S, 1995) in 3D interfaces, will this argument stand in AR environment? If so, how can we investigate the effect of 3D sound? For example, what measures can be defined to quantify the performance of depth judgments?
4. Normally the space of AR environments is relatively small, and the attenuation of sound with distance in this small area can be too vague to be perceived. This is because in most 3D sound systems, the intensity of sound (a very strong distance cue) would follow incremental reductions of 6 DB for each successive position of doubled distance (Begault, 1994). This means that for small areas, the 3D sound hardly plays a role to provide a cue to the depth (or the distance of the frontal sound sources). In order to make 3D sound give a cue to depth in a small space, we need to exaggerate the intensity difference of 3D sound according to distance (depth) changes of the virtual sound sources. However, this leads to a new issue, which is to research if this exaggeration will ruin the realistic and immersive feeling which we have been focused on. It needs to be investigated what are the subjective feelings on this exaggeration to the human.
5. Does 3D sound still improve the task performances in AR environment? If so, what kind of tasks can be used to evaluate a user's performance and what measures can be applied to investigate the 3D sound's effect on improving task performance?
6. Does 3D sound play a significant role in collaborative AR environment? If so, how can we apply 3D sound and evaluate its role? In what ways will the subjective experience 'human presence and collaboration' in AR environments? Especially what will be the subjective experiences from the addition of 3D sound?

Recent software and hardware advances in AR and auditory interfaces enable us to integrate 3D sound into AR environments and study these interesting questions. Although the quality of 3D sound generated by software synthesizers (Loki Entertainment Software)

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 003

is still to be examined by rigorous tests (Langendijk and Bronkhorst, 2000), the pre-tests taken by the developer and the user feedback from commercial applications have to some extend proved its fidelity. Our goal of this research is to do both quantitative and qualitative studies on the effectiveness of such software synthesized 3D sound in an AR environment. The outcomes of this research can have a useful impact on the development of AAR interfaces and the enhancement of AR environments for performing collaborative tasks such as co-operative teaching, training and gaming systems. Our contributions to the research topic are as follows:

1. *Depth perception in AR*: We investigated the impact of 3D sound on improving depth perception in AR systems with single-camera settings. We designed an experiment to measure the accuracy of depth judgment in the cases with and without 3D sound.
2. *Perceptual feeling of sound depth exaggeration*: In order to make full use of 3D sound's dimensional information in a relative small AR environment we exaggerated the intensity difference of virtual sound sources in different depths by multiplying an additional scalar factor. We investigated the subject's perceptual feeling to this exaggeration.
3. *Task analysis with AR sound*: We designed a searching task in another AR environment which requests the user to find a virtual character in the augmented scene; we investigated the impact of 3D sound on the task completion time.
4. *Collaboration and location with AR sound*: We applied a constant 3D sound on subject and task target to investigate the impact of 3D sound on the subject's feeling of 'human presence and collaboration' and the impact on 'locating correct augmented information'.

In Section 2, we discuss the general issues of developing AAR with 3D sound, and two different tracking methods for registering virtual 3D sound sources. The experimental goals and subjects are introduced in Section 3. The results of the experiments are then given and discussed in Section 4. The conclusions of the study are summarized in Section 5.

2. Development of AR environments with 3D sound for experimental analysis

2.1. General issues

Generating real 3D sound requires requires perceptual externalization of the sound source, which can only be brought about by using HRTFs (Bill G, 1994), reflections, reverberation, etc. (Wenzel, 1994) These HRTFs can be used to produce 3D sound effects from a linear convolution between sound signals and its corresponding spatial related impulse response (Wenzel et al., 1988) while the reflections and reverberations involve sophisticated modelling of environment. The individually derived HRTFs are better for individual users to provide more realistic sound experience. However, individual HRTFs are difficult for AR implementation due to the fact that AR systems are designed to be real time interactive systems that do not need individual user configuration. Thus our experiments were taken for randomized selected students. For easy implementation of 3D sound generation, several open source software development kits (SDKs), have been

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 004

developed such as OpenAL (Loki Entertainment Software) and FMOD (L. Firelight Technologies). These use the general HRTFs database and modeling environment in a relative simple manner. However, as mentioned in Section 1, for AR systems 3D sound does not play a significant role to tell the depth (or the distance of the frontal sound sources) because most of actual AR spaces are relatively of small size (for example small office, tabletops, books, etc.).

Sound intensity has been shown to be a key cue in distance perception in humans. However, in the small physical spaces used in AR systems the human is only effectively provided with binaural cues from the synthesizer. Thus we try to also add the effect of intensity for the depth cue. Our approach is to exaggerate the intensity difference of sound sources in different depths by applying an additional scaling factor to the intensity of 3D sound generated by synthesizer. The intensity of the ith 3D sound V_i' is given by Eq. (1):

$$V_i' = (D_{\max} - D_i)V_i/D_{\max} \tag{1}$$

where D_i is the depth of the ith 3D sound source; V_i is the intensity of the ith 3D sound generated by synthesizer given the position of the ith sound source. $D_{\max}$ is the possible maximum distance of the sound sources which is subject to the space of different AR environments. We measured the space in our first and second experiments and the maximum distance was 1.5 and 5.5 m, respectively.

Apart from intensity, it is known that for distance perception the ratio of the energies of direct and reflected sound is of importance (Middlebrooks and Green, 1991; Bronkhorst and Houtgast, 1999). For objects farther away from the listener, coupling the decrease in virtual sound intensity with an increase in the 'reverberation' will lead to a more realistic perceptions. In order to focus on the main points with regards to AR sound we did not add the factor of reverberation, and left this issue for our future research.

In our first experiment the sound sources are always placed around 1.5 m away in the front of the listener. However, in a more general and complex AR environment like the second experimental environment, the listener can walk freely in a physical space where virtual sound sources can be very close to the left and the right side of the head. In this case, the head starts functioning as a sound barrier for high frequencies. Consequently, the (high-frequency) sound intensity will of course increase as the source approaches the head in the ear at the same side of the head but, due to the increasing barrier function, decrease in the ear at the other side of the head. It is reported that binaural cues play an important role in auditory distance perception for nearby sources (Brungart and Rabinowitz, 1999; Brungart, 1999). This fact plus our exaggeration process mentioned above causes a design problem of AAR when the listener is very close to or even overlaid on the virtual sound source. In this case the user can have very strange auditory experience and even get confused. This was confirmed in our experimental results which are shown below (Question 3 of Experiment 2 below).

2.2. Development of AR environments with 3D sound for experimental analysis

In the most common AR paradigm, the user wears a head mounted display (HMD) through which the user views the real world, either directly, or via a video camera attached

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 005

to the front of the display Virtual objects are then superimposed on the real world. Early studies such as (Feiner et al., 1993) of this type simply overlaid two-dimensional text information on objects in the scene, but it has now become common to superimpose objects in three dimensions. The minimal requirements for this are that the position of the camera must be determined relative to the whole scene (or an object therein). The transformation matrix (TCM) between the scene and camera co-ordinates can then be used to generate the correct and seamless view of the virtual object. Given this TCM we can get the exact 3D position of the virtual object (virtual sound source) in the user's head co-ordinate system, which is the basic requirement of 3D sound synthesizer.

In our experiments, we applied the OpenAL (Loki Entertainment Software) to generate the 3D sound. Essential parameters needed are the virtual sound source's 3D position and its speed relative to the user's ear. More generally, the 3D position and speed of the virtual sound source and the position and orientation of user's head are needed. In the following parts we will introduce two tracking methods to solve this problem. However, since in our experiments the relative speed of head to virtual sound source is only as slow as 2 m s^{-1}, we ignored the Doppler Effect caused by their relative movement and thus we were only interested in their positions and orientations.

In our experiments, the final augmented virtual image was viewed through a Daeyang Cy-Visor DH-4400VP HMD, which presented the same 640×480 pixel image to both eyes. A PremaCam SCM series color security camera was attached to the front of this HMD. Our visual augmentation produced 25 images of 320×480 pixel resolution per second. To synchronize with the update of the visual scene, we updated the position of the virtual sound sources 25 times per second. The headphone on the HMD plays back the 3D sound generated by the computer. In this way, we provide an AR environment which is both visually and audibly augmented.

2.3. AR experimental environment implemented by vision-based tracking method

There have been many vision tracking methods for AR reported in recent years. In our first experiment, we used a precise vision-based toolkit for AR tracking (Kato and Billinghurst, 1999) reported by Kato and Billinghurst. It used image-processing technology to retrieve 3D position information from a fiducial marker. The 'ARGroove' (Poupyrev et al., 2000) applied this tracking method to get different position and orientation descriptions of the marker movements for which they defined different music. Our first AR experiment system used this tracking method to track the position of virtual sound source which is attached on fiducial marker.

The 3D position of the marker relative to camera can be obtained by computing the TCM from the camera co-ordinates to the 7 marker co-ordinates (Kato and Billinghurst, 1999). However, the 3D position of the marker we obtain is based on the camera co-ordinates. Thus, in order to obtain the 3D position of the marker in user's head co-ordinates (3D sound synthesizer's co-ordinates), simple translations from the camera to the center of user's inter-aural axis should be considered. As shown in Fig. 1, in order to attach the virtual sound source on the marker, the TCM from the 3D sound co-ordinates to marker co-ordinates could be calculated by simply applying translations to TCM obtained before. The 3D position of the marker in the 3D sound co-ordinates are then fed

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 006

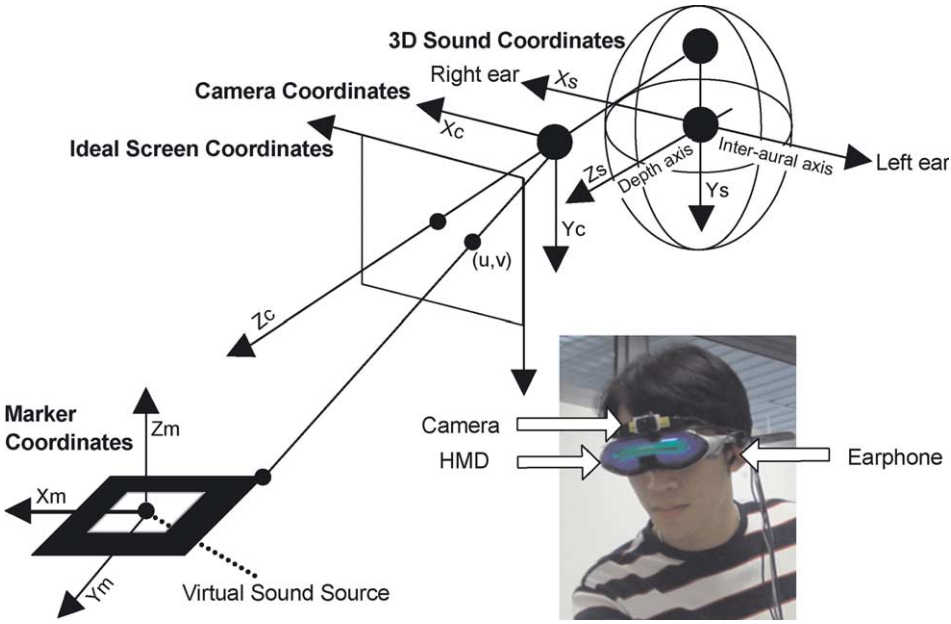

Fig. 1. In order to place a virtual object in the image of the three dimensional scene, the Euclidean transformation between the camera co-ordinates (X_c, Y_c, Z_c) and the marker co-ordinates (X_m, Y_m, Z_m) is estimated. This calculation is based on the imaged position of the corners of the square marker. The imaged position of one of these corners is depicted with pixel co-ordinates (u,v). To obtain the transformation matrix from the 3D sound co-ordinates to marker co-ordinates, translations should be taken from the camera co-ordinates to the 3D sound co-ordinates (X_s,Y_s,Z_s). The picture at the right bottom shows the physical setting of our single camera system where camera is mounted to the front of HMD and 3D sound plays back into the earphones.

into OpenAL (Loki Entertainment Software) for synthesis. Therefore, the 3D sound appears to emanate from the marker.

2.3.1. Depth confusion problem in single camera AR environment

In the most common AR paradigm, a single camera is used to capture the physical scenes (Billinghurst et al., 1998; Zhou et al., 2002; Kato and Billinghurst, 1999) This is not only because this setup is cheap but also because that it requires little computation cost and can be easily implemented. However, although this kind of AR environment setup enables the first-personal viewpoint, it cannot provide stereo video images for each eye, and this can lead to the difficulty of depth perception. Solving occlusion problems (Wloka and Anderson, 1995) or using stereo cameras can improve the depth perception problem. However, it requires more sophisticated settings (Stockburger, 1998) and more computational cost (Kanbara et al., 2000; Takagi et al., 2000). Thus, the low-cost single camera mono system is still widely used for AR applications.

Our objective is to find out whether 3D sounds could be helpful to improve depth perception in this commonly used lower-cost mono AR environment. Fig. 2 shows the depth confusion problem related to this kind of AR environment setup. We know that

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 007

Fig. 2. The depth confusion problem in a single-camera AR setting. The left column shows the image sequences of a scene captured by a moving camera while the right column shows the augmented view seen by the user. (1) In the top left frame, two markers of different size are located in different depth which is hard to tell because of lacking cues. In the following frames, we know that the left marker is nearer to the camera with the cues obtained from the motion, background and lighting environment. (2) Right column shows that the augmented objects occlude everything behind them hence weaken the effect of motion cue. Without background information (in this special case, the objects are located on the same board), relative depth is hard to tell.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 008

human depth perception is based on both monocular cues and binocular cues. Binocular cues include parallax and ocular convergence, while monocular cues include interposition, relative (or familiar) size, linear perspective, texture gradient, light and shadows. The users' responses in our user study suggest that the depth of augmented object is hard to decide in this kind of mono AR environment setup.

This is because that the binocular cues do not work well due to the single-camera setting. Furthermore, because the augmented objects were rendered over the captured images, they appear to occlude everything behind them. To render the virtual objects to be faithful to the actual lighting conditions of the physical world is still a challenging problem. The correct display of shadow and texture gradient of the augmented objects needs the precise estimation of light sources (Kanbara and Yokoya, 2002). The occlusion problem of the augmented objects relative to the real object is also a difficult problem (Fuhrmann et al., 1999). Due to these difficulties, it is still not possible in real-time AR systems to have realistic virtual objects with respect to lighting, shadowing, and occlusion. Thus, our goal is to find out whether 3D sound can help in the judgment of depth, and if so, to what extend can it be helpful for AR.

We thus designed an experimental task which requires the subject to tell which of the augmented objects is closer (less deep) to her and we calculated the accuracy of the judgments. The subjects were required to give their judgment under the conditions of vision plus 3D sound and vision only. As will be seen in the sections below, the judgment accuracy was found to be significantly higher when 3D sound was applied.

2.4. AR experimental environment implemented by acoustic-inertial tracking method

The vision-based maker tracking method can register the virtual objects and virtual sound sources into the real scene precisely. However it has several short-comings as well. For example, the marker tracking algorithm is sensitive to the lighting environment. Also the requirement of keeping a marker in the scene all the time limits the user's movement as well as viewpoint and thus restricts the user to a relatively small space and a narrow angle of view. This drawback could be improved to some extent by using multi-marker tracking or natural feature tracking methods (Neumann and You, 1999). However, they still do not allow fully free head and body movements.

Some previous work used other methods for registering a virtual sound source in a larger space which allowed user's free movement. Makino et al. (1996) used the Global Positioning System (GPS) to locate the traveler within the environment. With differential correction by way of audio link from a GPS reference receiver in the local vicinity, localization is accurate to around 1 m under good conditions of reception. The listen project (Eckel et al., 2001) used motion-tracked wireless headphones to retrieve the position of the user in the environment. A similar method was applied by Schmandt (1998) where the head position is gained from a Polhemus sensor mounted on the headphones.

In our research system (Cheok et al., 2004; Wang et al., 2002), we use the InterSense IS-900 system (InterSense) for precise motion tracking. The IS-900 system is a hybrid acoustic-inertial six degree of freedom (6 DOF) position and orientation tracking system.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 009

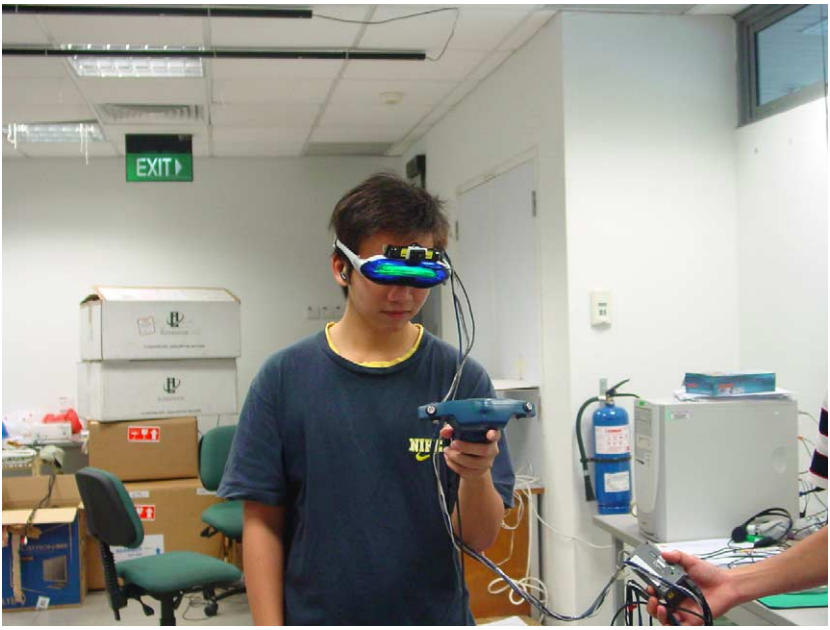

Fig. 3. The physical set-up of our acoustic-inertial tracking AR environment. On the ceiling there are six IS-900 SoniStrips which makes up a three meter by three meter area. A head tracker is mounted on the user's HMD and a camera is fixed on the head tracker. A wand tracker with four buttons and one joystick was held in the user's hand as an interaction device.

It tracks changes in orientation and position by integrating the outputs of its gyroscopes and accelerometers, and corrects drift by using a room-referenced ultrasonic time-of-flight range measuring system. It can obtain a resolution for position of 15 mm.

For the experimental system, as shown in Fig. 3, we mounted six IS-900 SoniStrips on the ceiling of a 3 meter by 3 meter space. This system can support up to four tracking stations. We used four tracking stations to track two users' heads and hands, respectively. A head tracker was mounted on the user's HMD and a camera is fixed on the head tracker. A wand tracker with four buttons and one joystick was held in the user's hand as an interaction device.

The system superimposes graphics on video sequences captured from the camera and displays the combined images on the HMD. We employed OpenAL (Loki Entertainment Software) for generating 3D sounds. Unlike the vision-based tracking method where we make the user's co-ordinates system as the reference, in this method, we used the IS-900 co-ordinates system as a reference while using the users' head position, orientation, and the virtual sound sources' 3D position as input parameters for the OpenAL software.

3. Experimental goals and subjects

Two experiments were taken in the two pre-mentioned AR environments, respectively. In the first experiment, our goal is to find out:

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 010

1. Whether 3D sound helps depth perception and by how much.
2. Whether the exaggeration of intensity difference of 3D sound will be acceptable or unacceptable psychologically.

In the second experiment, our goal is to find out:

1. Whether 3D sound helps in task performance, or more specifically, whether it helps to shorten a searching task completion time and by how much.
2. Whether 3D sound gives more immersive and realistic perception to the users and whether it helps the subject to distinguish different objects in AR environments.
3. Whether 3D sound increases the feeling of 'human presence and collaboration'.

Both qualitative and quantitative studies were conducted in each experiment. The study took place over a four-week period in the summer of 2002.

Forty subjects (13 females and 27 males) were each paid Singapore $8 (approximately $4 USD) per hour to participate in this study. All the participants were undergraduate students at the National University of Singapore. Most of them were freshmen or juniors majoring in computer engineering. The subjects were on average 21.3 years old and had used computer for 8.7 years. They were asked to mark from 1 to 7 about how often (1 means very often while 7 means very seldom) do they use computers and how often do they play games. The averages were 1.7 and 4.5 respectively, which means that they often use computers, but play games only occasionally. All of the subjects reported clear-vision and normal hearing abilities.

Each subject performed both experiments. However in the second experiment, 20 subjects (seven females and thirteen males) were asked to conduct the searching task with the aid of 3D sound while the other 20 subjects (6 females and 14 males) conducted the same task without the 3D sound. Before doing the experiments, all the subjects were given a brief introduction about AR and the tasks to be performed. Each of the subjects was given five minutes to initially try on the HMD and experience the 3D sound effect. After the experiments, all subjects were given 25 min to finish a questionnaire based on the experiments finished.

4. Experimental results

In this section we present the details of the evaluation studies of 3D sound effect and performance in our two AR environments described before

4.1. Experiment of the impact of 3D sound on depth judgment

4.1.1. Design and procedures

In this experiment, we examined the effect of 3D sound on helping users to judge the relative depth of objects in an AR environment (Fig. 4) implemented by vision-based tracking method. Users were required to tell the relative depth of the two markers upon which virtual objects were augmented. In order to precisely position the markers, we used a white board to hold the markers. We set the height of user's chair so that the markers

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 011

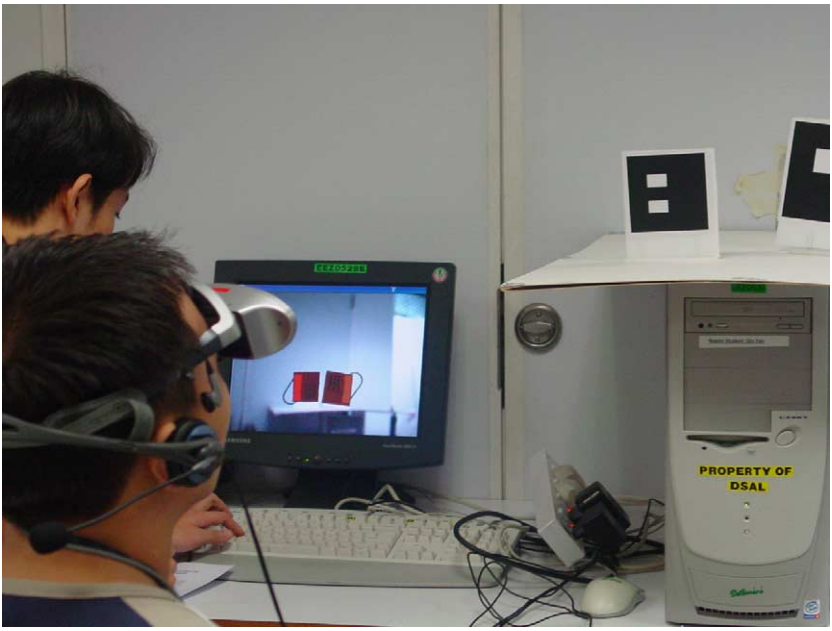

Fig. 4. The physical set-up of the experiment on depth judgment. The output of the computer VGA signal is send to a monitor and HMD for display by a video converter. The user can only see from HMD while the monitor is for the use of the experimentalist. Two virtual telephones with different colors are augmented onto different markers respectively. The size of augmented objects is adjusted to about the same by adjusting the position of marker. Users are required to tell the relative depth of the two objects in the cases of with and without 3D sound aided.

and the eyes were at the same height. By this way we eliminated the vertical vision cue to depth perception-as the relative position on the same plane can tell the relative depth. This makes sense in AR environment since the augmented objects are not necessarily located on the same plane, and thus this cue is not commonly available. The relative position of the two markers was chosen carefully so that the two markers seem to be approximately of the same size in the camera's view. As described in Section 2.3.1, this will cause depth confusion and hence we can carefully decide the effectiveness of adding 3D sound to the virtual objects.

A preliminary study (Mereu, 1995) indicated that visually impaired users can use the sound environments to perceive depth and position in a 3D application. We hence propose that 3D sound could also be helpful in AR environment for people to perceive depth. To test this hypothesis, we implemented 3D sound in our AR environment and measured the accuracy of relative depth judgment of subject.

4.1.2. User interfaces

As shown in Fig. 4, we augmented the same shape virtual telephone with a different color on each of the markers. The virtual telephones were augmented with a 3D telephone sound that was produced to emanate from each virtual telephone. As explained in

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 012

Section 2, the intensity difference of 3D sound due to different depth of markers was exaggerated by a scaling factor. Subjects were asked to put on the headphone of HMD and perceive the difference of 3D sound when the marker was changed to different positions.

Every subject was required to perform 10 trials judging the relative depth of two telephones. Before every trial, the signal of the HMD is turned off so that the subject could not see how the markers were arranged. During that time, the user only sees black because the HMD is a nonoptical see through type. We placed the two different sized markers on the condition that the image size of the two objects are of about the same size, this can be observed by a monitor which is connected to the same VGA output as what the HMD is connected. Then the HMD is turned on and the subject was required to tell which telephone was nearer to him in the cases with and without 3D sound aided respectively. The possible answers are 'left', 'right' and 'uncertain'. We recorded the subjects' judgments and compared them with the correct answers for numerical analysis. We also asked some questions in the questionnaire for psychological analysis purpose.

4.1.3. Results

4.1.3.1. Performance.. The chart in Fig. 5 provides a visual summary of the results of the depth judgment experiment. Fig. 5 compares the number of correct relative depth judgments in 10 trials for 40 subjects in the cases with 3D sound and without 3D sound aided respectively. It can be seen that judgments with 3D sound aided lead to 2.5 times as accurate (70.25 versus 28.25%) as those without 3D sound added.

Our expectation of this experiment was to evaluate the impact of 3D sound on depth judging accuracy. Using a 0.05 significant level, a one-way ANalysis Of VAriance (ANOVA) (Stockburger D, 1998) of the individual correct judgment times of each user was used. In this experiment, we obtained $F(1,39) = 129.55$; $p < 0.001$ which declares the result to be statistically significant. Thus we conclude that 3D sound does significantly help to increase the number of times of correct depth judgments.

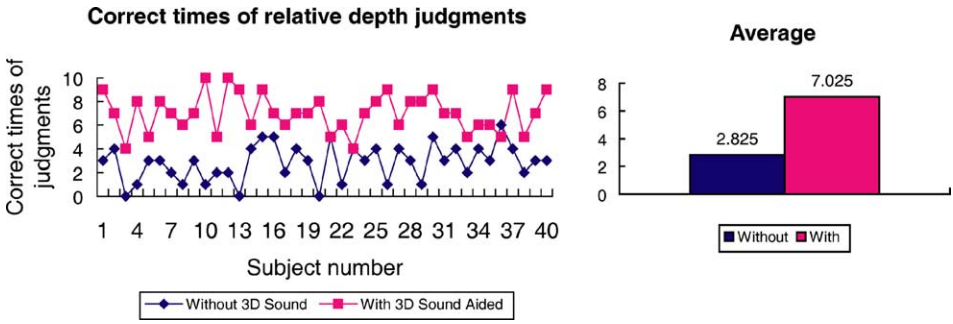

Fig. 5. Comparison of the correct depth judgment times of 40 subjects in ten trials in the cases of with and without 3D sound aided. From this chart we can see obviously that the accurate judgment times increase drastically by introducing 3D sound. The average of correct judgment times is increased from 2.825 to 7.025 by introducing 3D sound.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 013

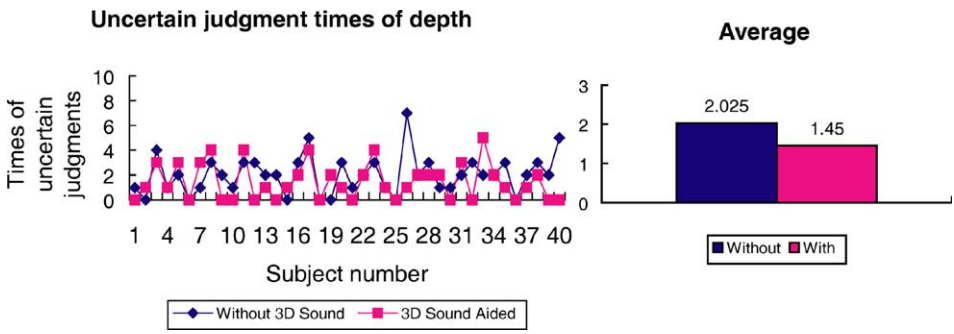

Fig. 6. Comparison of the uncertain judgments in the cases of with and without 3D sound aided. The average of uncertain judgment times is decreased from 2.025 to 1.45 by introducing 3D sound.

On the other hand, the addition of 3D sound decreases the uncertainty of relative depth judgment. Fig. 6 shows the comparison of uncertain judgment times between the cases of with and without 3D sound aided. The addition of 3D sound decreases the average uncertainty from $M=2.03$ to 1.45 times. A significant difference is found ($F(1,39)=4.24$; $p<0.05$). These results proved that the depth perception was significantly enhanced by 3D sound.

There was no significant difference found between the genders. For the correct judging times without 3D sound, males have a mean value of 2.67, females have a mean value of 3.15, $F(1,38)=0.9$; $p>0.1$. For the correct judging times with 3D sound aided, males have a mean value of 7.00, females have a mean value of 7:08, $F(1,38)=0:02$, $p>0.1$.

4.1.3.2. Questionnaire responses.. The purpose of the questionnaire is to examine how the subject qualitatively judges the relative depth of augmented objects and how difficult the task is. Apart from the quantitative results we obtain from the above experiments, it is advantageous to know the subjective feeling of users due to the addition of 3D sound, and the process of the sound intensity exaggeration. The main questions in questionnaire about this experiment were as follows:

1. When you were supposed to judge the relative depth of the two markers, what cues did you mainly use the depth of markers (multiple choices are possible)?
   A. brightness B. color C. size D. intuition E. background F. motion G. others
2. Do you think it is hard to tell the relative depth of two objects in this experiment without 3D sound? Please select your answer from the following choices:
   A. very hard B. hard C. medium D. easy E. very easy
3. Overall, do you think 3D sound helps you to make a correct decision?
   A. yes B. no C. uncertain
4. In real life situations, the intensity of sound doesn't change drastically when the sound source is moving within a short distance, say 100 mm. However, in our experiment we exaggerated this effect. Thus you perceived an obvious difference between the two markers when they were placed in a relative small different depth.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 014

Do you feel it is unintuitive that the depth scale of 3D sound in this experiment has been exaggerated? Please select from the following choices about how you feel:
A. very unintuitive B. unintuitive C. medium D. intuitive E. very intuitive F. did not notice at all

Question 1 allows multiple choices. Every choice given by the user will be counted as a single vote to the selected choice. Although different users might tick different numbers of choices, the vote diagram is not severely disturb qualitatively since our purpose is to find out what are the cues most users will use for judging depth. As shown in Fig. 7(a), the 'size' cue was given the first place (30 votes) of all the cues in our question. This partially explains why

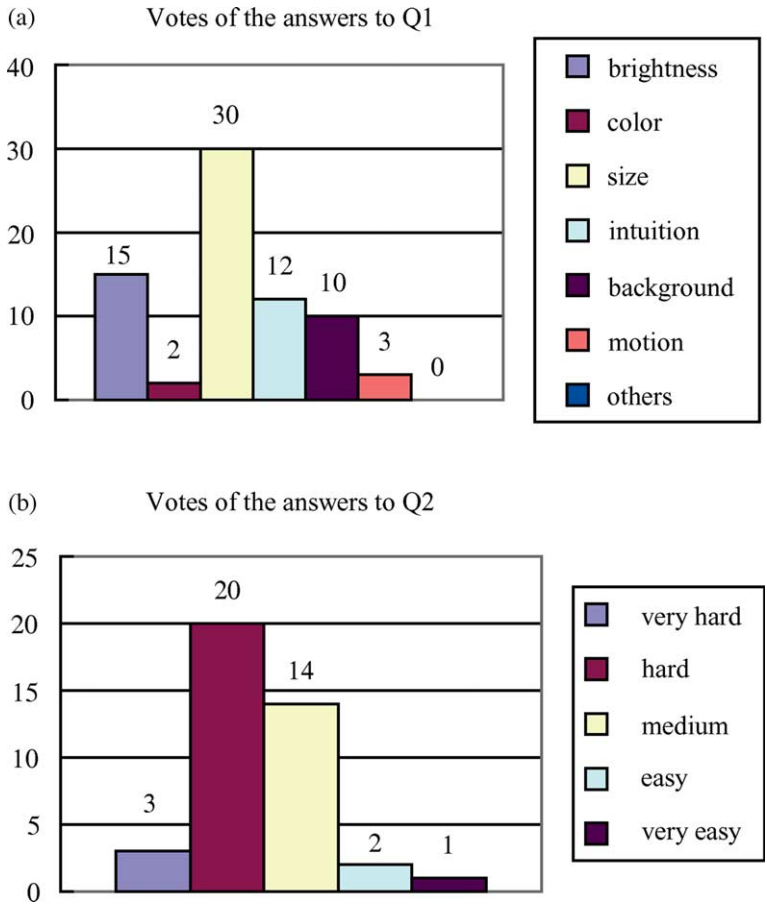

Fig. 7. Statistical charts of the answers to Question 1, 2 of Section 4.1.3 (a) The votes of the answers to question 1: When you were supposed to judge the relative depth of the two markers, what cues did you mainly use the depth of markers (can have multiple choices)? It can be seen from this chart that users are more likely to judge the depth of object by the size, brightness and the background cues. (b) The votes of the answers to question 2: Do you think it is hard to tell the relative depth of 2 objects in this experiment? It can be seen from this chart that it is hard to tell the relative depth of the two markers in our experimental AR environment.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 015

it is difficult for subjects to judge the relative depth when two markers are almost of the same size. Our setting of AR environment in this experiment restrained the effect of background (ten votes) and motion (three votes). The cue 'brightness' which reects the cue 'light and texture gradient' obtains the second highest votes (15 votes). This gives us a hint that in order to obtain good depth perception, it is also important to adjust the appearance of virtual object so that it fits its surrounding physical lighting environment.

However, in this experimental system, the brightness of the object is predefined when the model was created. Hence it brings more difficulties for depth judgment, and it can not provide the cue 'light and texture gradient'. Therefore, this suggests that in the future research we can adjust the brightness of augmented objects according to its lighting environment so as to achieve better depth perception.

In response to question 2, most subjects claimed that it is hard (three claimed very hard and 20 claimed hard) to judge the relative depth of objects in this experiment without the 3D sound cues. Fig. 7(b) gives the chart of this result. This psychological feeling is in accordance with the results shown in Fig. 5 where the accuracy of judgment is only 28.25% in average for tests without sound.

There is an interesting result from the answers to question 3. Only 19 subjects thought that 3D sound has helped them however, 6 subjects claimed 'no' and 15 subjects claimed 'uncertain', as shown in Fig. 8(a) (one subject somehow forgot to give the answer to this question, thus only 39 votes are appeared in this chart). This implies that the addition of 3D sound may have contributed to the depth perception (Fig. 5) although the subject may not psychologically feel it. While this result is quite interesting, more detailed and long term psychological investigation needs to be conducted before drawing any reliable conclusions.

In question 4, our prospect is to find out whether our exaggeration of 3D sound will lead to a degrading of the perceptual experience. The result (Fig. 8(b)) shows that only one subject claimed that it is 'very unintuitive' and six subjects claimed that it is 'unintuitive'. Nine subjects claimed 'medium', 12 subjects claimed 'intuitive' and one subject claimed 'very intuitive'. Moreover, 11 subjects claimed that they 'didn't notice at all'.

Although there is still a risk of losing the 'realistic experience' since the 3D sound environment is different from our everyday life, this result suggests that most of the users will still feel the virtual scene is intuitive psychologically if the exaggeration scale is properly chosen. This result is critical to the development of 3D sound system in AR environment since, unlike in pure VR system, normally the exaggeration is required, otherwise we may lose the information carried by one dimension of 3D sound-the depth. This occurs especially when the physical AR environment is as small as our experimental game system or even smaller, as in the desktop AR environment of the depth perception experiment.

4.2. Experiment of 3D sound's impact on task completion time and the feeling of 'human presence and collaboration'

4.2.1. Design and procedures

In the second experiment, we examined the effect of 3D sound in a larger AR environment in the form of a simple game Barfield (Barfield et al., 1997) identified three primary benefits of auditory spatial information displays as: relieving processing demands

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 016

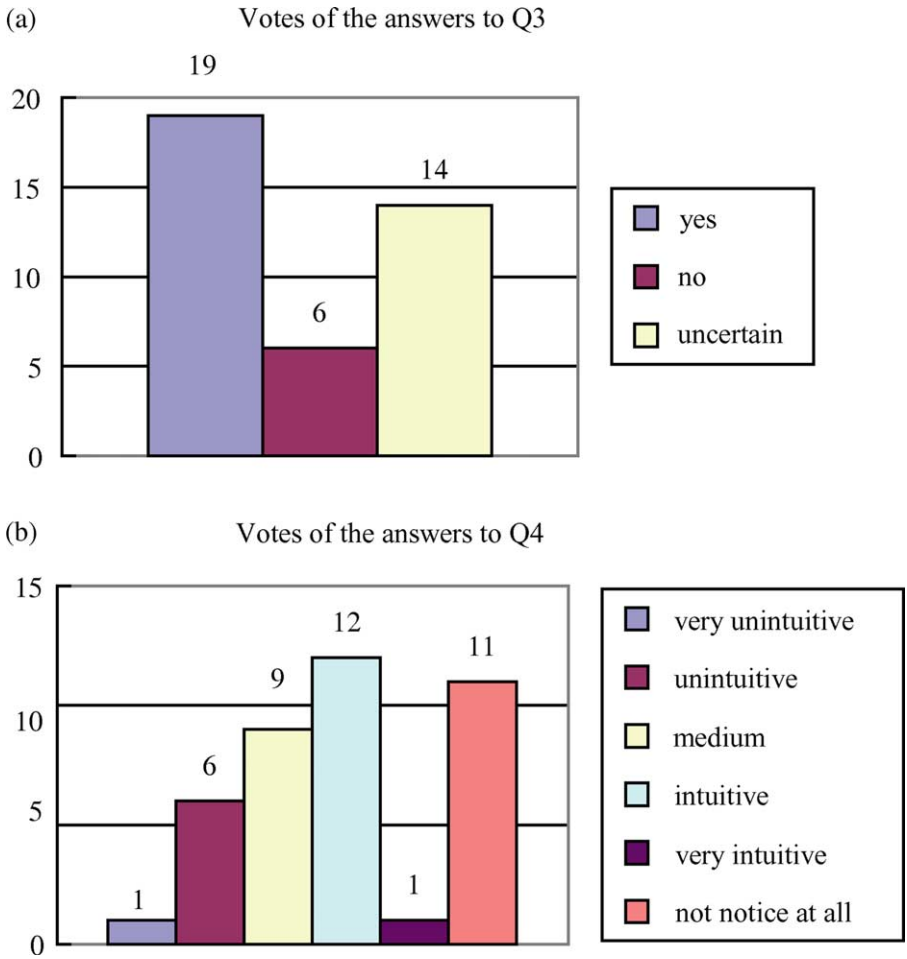

Fig. 8. Statistical charts of the answers to Question 3, 4 of Section 4.1.3. (a) The votes of the answers to question 3: Overall, do you think 3D sound help you to make a correct decision? (One subject somehow forgot to give answer to this question, thus only 39 votes are appeared in this chart). (b) The votes of the answers to question 4: In our every life, the intensity of sound doesn't change drastically when the sound source is moving within a short distance, say 100 mm. However, in our experiment we exaggerated this effect. Thus you perceived an obvious difference between the two markers when they were placed in a relative small different depth. Do you feel unintuitive that the depth scale of 3D sound in this experiment has been exaggerated?

on the visual modality, enhancing spatial awareness, and directing attention to important spatial events In our research system, we extend the benefits in this experiment by:

1. helping to localize an object in AR environment when the visual modality is weakened;
2. enhancing the feeling of human collaboration;
3. directing attention to different spatial objects or helping user to identify different spatial objects.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 017

4.2.2. User interface

Our experimental system allows two users to play a simple game at the same time The objective of this game is to rescue the princess who was captured by an evil witch. In this experimental game, three stages were used to for investigating the effects of 3D sound. In the first stage, users are allowed to walk around in a three by three meter region and to explore a virtual land augmented on the physical ground through a small rectangle virtual window displayed on the tracked wand. Fig. 9 shows the scrolling 3D landscape when the player is moving her hand. A small virtual plane is attached to the wand as an avatar for the user, as if the user is exploring the game space by manipulating the plane. The player is looking for the castle through the 'magic 3D window'. Two 3D sounds are applied in this stage and can be heard by both players. The first one is a 3D sound 'Help!' which was applied to the princess located inside the castle. The second sound is a constant buzz applied to the wand tracker of co-player as if it is generated by the engine of co-player's airplane. Each player can hypothetically localize his/her co-player by localizing the buzz sound.

In the first stage the player is allowed to become more and more familiar with the searching task and the perception of 3D sound. We purposely design this stage as a training process. When the player walks closer to the sound sources (the castle and the princess therein or the co-player), the sound becomes louder while it becomes softer

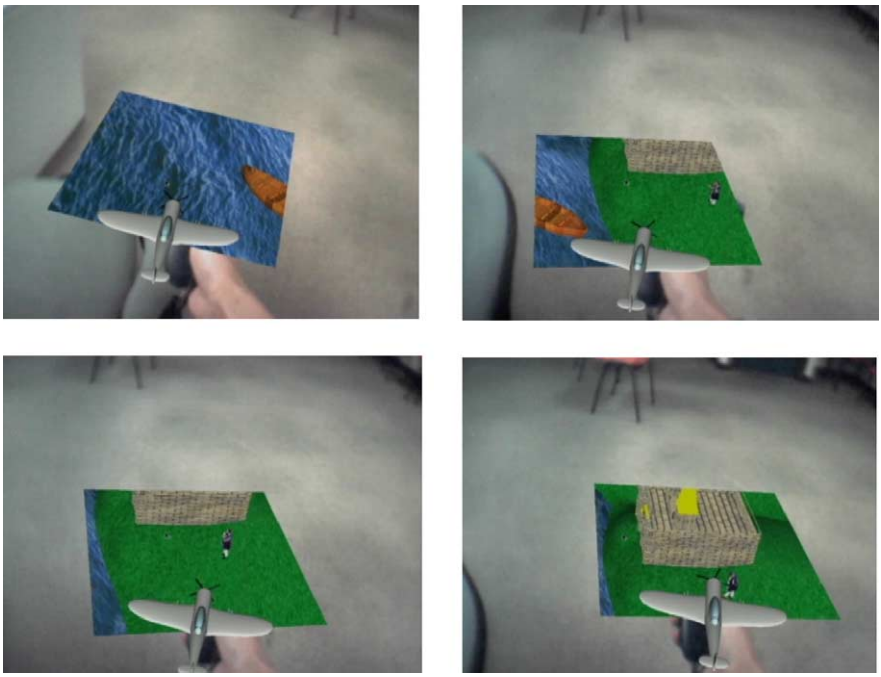

Fig. 9. Looking for the castle through a 'magic 3D window'. A virtual plane is augmented on the wand as an avatar for the user. A rectangle window is augmented in front of the wand through which the user can see the virtual land on the ground. When the user is moving around, the 3D landscape is also scrolled in the 'magic 3D window'. The user is looking down at the castle as if she is looking down from the sky in the airplane.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 018

when the player walks away from the castle. An exaggeration of the intensity difference described in Section 2 was undertaken here as well. Furthermore, it should be noted, when the player rotates her head or body, she can perceive that the sound source is also moving relative to her movements. We updated the position and orientation of the head tracker 25 times per second which allows a quite smooth feeling of sound source movement.

After the castle is found, players move to the second stage. They battle with a virtual witch who has captured the princess. The virtual witch is augmented in the physical world and is hovering in the experimental region. Players can move freely in the physical space to get the best shooting angle and position. They need to reload or transfer their bullets by putting their wands together which encourages the collaboration in the game, as shown in first two frames of Fig. 10. Apart from the sounds used in stage one, we applied the other two 3D sounds, the cry of witch when it is hit and the explosion sound when it is exploded after being shot for 10 times.

After destroying the witch, the players can fly into the virtual scene and rescue the princess from the castle. This enables a seamless transition from AR environment to VR environment. After the user flies into the virtual scene, she enters a virtual environment in which players can use joystick on the wand to move around, while they can turn over their head or body to change the direction of viewpoint. In this stage we still keep the princess's cry 'help' but we change the sound of co-player from a buzz of engine into a footstep sound. Both of the players can be either in AR or VR mode and can continuously be aware of each other by both vision and sound. Fig. 11 shows the views in this stage.

To examine the impact of 3D sound on task completion time and subjective feelings in this AR environment, 3D sounds were applied in this experimental game. Their functions are summarized as follows:

1. *Players' sounds*: A constant buzz sound is applied to the co-player's airplane when the users are exploring the augmented land. Since the plane is attached to the wand, the 3D positions can be easily obtained from the tracker in the wand. By applying this sound, the user can always hear where is the location of the co-player, even when the co-player is not visually in the scene. She can still roughly tell to what direction is her co-player located and whether the co-player is getting closer or further to her. This buzzing sound is changed into a footstep sound when the co-player enters the VR environment. The different sound gives the hint of whether the co-player is in the AR or VR environment as well.
2. *Human voice*: A constant human voice 'Help! Help!' is applied to the princess which is the target of the searching task. This sound helps the player to turn and face to the correct direction and walk to the correct location where she can see the princess through the small window.
3. *Entertainment sounds*: To make the game more entertaining, we also applied other sound effects. A shooting sound is played when user press the shooting button on the wand. A scream sound and an explosion sound were played when the witch is destroyed. A landing sound of engine is played when the user transits from AR environment to VR environment. A female voice saying 'congratulations' and some greeting words was played when the user finished her mission-find the princess and rescue her.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 019

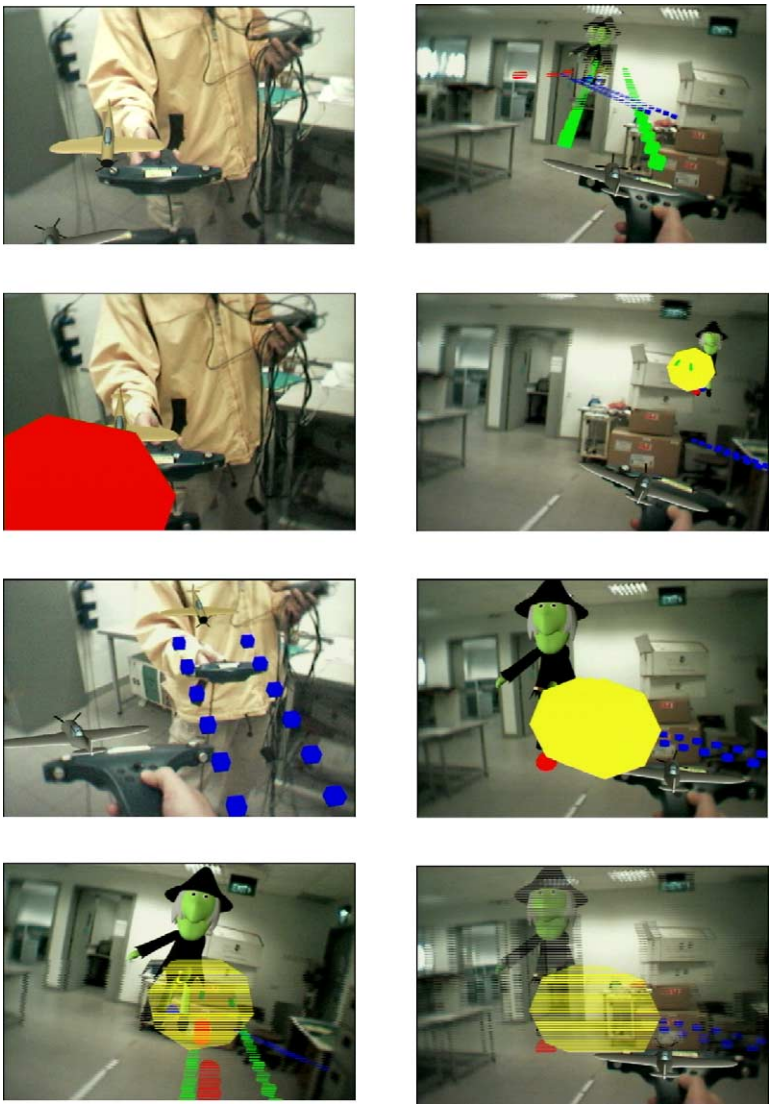

Fig. 10. Two users are collaboratively fighting the witch hovering around them in the physical experimental area. The user can see the augmented virtual witch in three dimensions through HMD. The two users need to reload or transfer their 'bullets' by putting their wands together as shown in the first two frames. When the user pressing a specific button on the wand, the airplane can fire bullets (blue and green dot lines in the subsequent frames). The position and orientation of the wand are used to decide trajectory of 'bullets' and whether the witch is shot or not. In the last frame, the witch disappeared after being shot for ten times.

In this experiment, we took both numerical and psychological studies of the subjects' responses. After the players finished the game missions, they probably have been familiar with the game, an additional task is required to perform in stage one. Instead of searching the castle, they need to search for the princess by walking around and scrolling

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 020

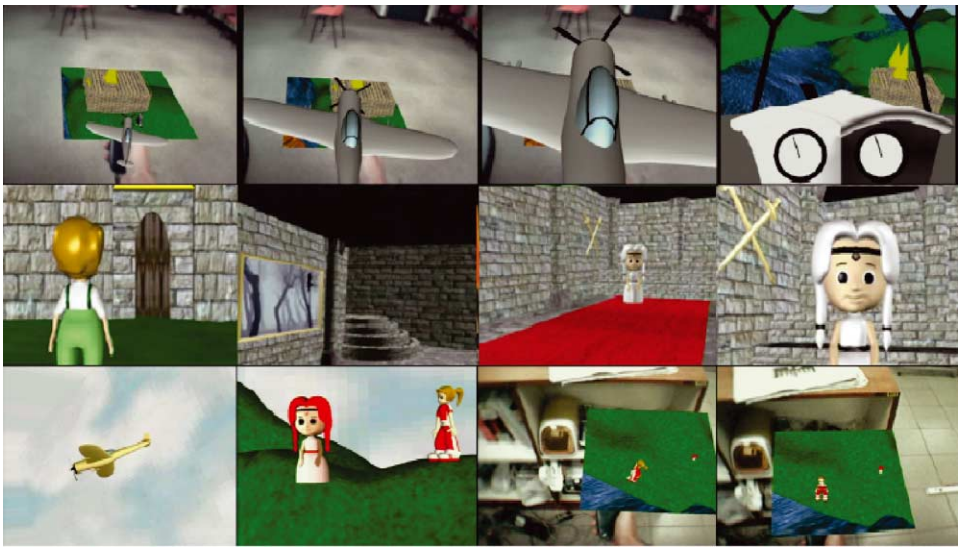

Fig. 11. Rescue the princess in the castle with co-player. (1) The first row shows the seamless transition from AR to VR that the user's airplane is 'flying into' and 'landing on' the VR world. (2) The second row shows one player's view when they are on the way to rescue the princess in the castle. The user uses the joystick to navigate however they can turn the body and head of the virtual character by physically turning their body and head. The footstep sound comes from the co-player and the 'Help!' scream of the princess can always be heard. (3) In the first frame of the third row, a user can physically look up in the sky to find her co-player who is still in AR mode (before flying into the VR mode). The user in the AR mode can still see the co-player and the princess by looking down to the land as if she is on the airplane. The second frame shows one player's view point when both of the players are in VR mode and come to the princess.

the window. The princess was taken outside the castle and was put at an open area on the land. Each subject was given a trial to see how the princess looks like in the small window. Then the princess was put at another position which is the same for all subjects. Every subject started from the same position and orientation in the space and began to search for the princess. The completion time of this searching task was recorded. Twenty subjects (7 females and 13 males) were asked to conduct the searching task with the aid of 'Help!' sound while the rest twenty subjects (6 females and 14 males) conducted the same task without any sound. Every subject performed this task only once either with or without sound aided. Apart from this numerical study, psychological studies were conducted by a linguistic questionnaire.

4.2.3. Results

4.2.3.1. Performance.. In this experiment the subjects performed a searching task as described in Section 4.2.2. The chart in Fig. 12 compares the completion time (in seconds) of the searching task for two groups in the cases with 3D sound and without 3D sound aided. The average task completion time of the group without 3D sound aided was 1.5 times as long (133.05 versus 88.1) as that with 3D sound aided. The effect of 3D sound on shorten the task completion time is significant, $F(1,38)=6.7$; $p<0.025$.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 021

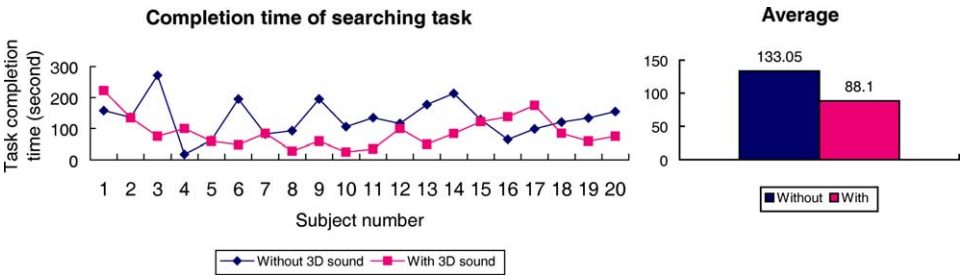

Fig. 12. The comparison of searching task completion time in the cases of with and without 3D sound aided. The average of completion time is shortened drastically from 133.05 to 88.1 s by introducing 3D sound.

No significant difference is found between the genders in this experiment as well. For the task completion time with 3D sound aided, males have a mean value of 88.4615, and females have a mean value of 87.43, $F(1,18)=0.0018$; $p>0.1$. For the task completion time without 3D sound aided, although females' mean value is higher than that of males' (166.33 versus 118.79), no significant difference is found at the significant level of 0.05, $F(1,18)=3.03$; $p>0.05$.

From the results given above we can see that, 3D sound helps significantly to shorten (about by half) the searching task completion time in this AR environment.

4.2.3.2. Questionnaire responses. The purpose of the questionnaire questions is to examine the impact of 3D sound on the subject's psychological feelings. We wanted to examine whether 3D sound improves the feeling of immersiveness and human presence and collaboration, whether it helps to identify spatial objects, and how much impact it has on these feeling. The main questions in questionnaire about this experiment were as follows:

1. Does 3D sound improve your perception of immersiveness in the AR environment? Do you think it is more realistic with 3D sound than without it? Why?
2. Does the 3D sound improve your feeling of human presence and collaboration as you can hear the sound coming from your collaborator at anytime?
3. Does the 3D sound help you to identify different information signals (spatial objects)?
4. Were you aware of your co-players in different places?
   A. always B. very often C. often D. barely E. never
5. What has helped your feeling of 'human presence and collaboration'? (can be multiple choices, please order the factors by the order of decreasing importance)
   A. small talk between co-players B. vision cues C. 3D Sound cues D. Others

According to question one, almost all subjects (39 of 40 subjects) claimed that 3D sound did improve the perception immersiveness. A total of 38 subjects claimed that it is more realistic with 3D sound, and one subject claimed 'a little better'. The main reasons they provided to support their viewpoint are as follows:

1. 'It is natural feeling and more realistic with 3D sound.'
2. 'You can feel the sound from all directions which gives a more 3D feeling.'

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 022

3. 'You can tell where the collaborator is positioned, or to know the surrounding objects in the AR environment.'
4. 'The feeling is closer to real world and 3D sound helps to navigate.'
5. 'I can hear and judge distance from the object producing the sound.'

From these responses, we can find that 3D sound does play an important role complementing the visual AR environment. Although this result seems to be natural since the same conclusion has been drawn in virtual reality environment (VRE), it proves that this conclusion is still tenable in AR environment when the implementation method has been changed and especially when the interfaces preserve our social interactions which are weak in VR.

There were also negative comments. One subject suggested that there should be more background sound like noises to make it more realistic. The subject said, 'The sound field is too quiet to be realistic.' In fact, this is consistent with the argument given by Begault (Begault, 1994): "Even with non-spatial cues, one can tell something about the distance of different sounds and their context,…, but the inclusion of 3-D sound techniques and spatial reverberation cues can greatly increase the immersiveness of a simulation, as well as the overall quality and the available nuance in representing different environmental contexts".

The important result we obtained from the responses to question two is that 3D sound improves the feeling of 'human presence and collaboration' when the subject is performing this experimental task. In this question, 39 of 40 subjects answered 'yes'. Some of the comments were:

1. 'At least I can find whoever it is.'
2. 'The sound constantly reminds us of the presence of the collaborator.'
3. 'Yes, like someone is actually in the world with you.'
4. 'It is really interesting that you can turn to face him when the clop-clop is getting closer to you in your back.'

One interesting comment was that 'I can know whether he is still in the airplane or has landed down and walking around' (which meant that the user could know if the other player was in AR or VR modes). This is actually one function of sound which is serving as an auditory icon (Gaver, 1986). A good negative comment suggested that 'it can be improved by adding speech communication, because a running motor does not sound very human, so it is like an object somewhat'.

Apart from applying 3D sound to the collaborator, we also applied 3D sound to the princess who is the target of the task. The responses to question three suggest that 3D sound, when applied to different spatial objects (information signals) in AR environment, can help the user to identify different information carriers. In our experiment 36 subjects claimed that they can constantly be aware of both the other collaborator and the princess simultaneously. This result suggests that by introducing more than one 3D virtual source sources in AR environment, it is possible for the users to identify different meanings of them and to distinguish them by the localization of the sound sources. As one subject described, 'yes, very much so. For instance, if the collaborator sounds from the left

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 023

earpiece and princess cries are heard at right earpiece, the directional news determines the positions of the collaborator and princess.'

However, one valuable criticism complained that 'the volume differs when you are far but when you are near I could not hear a difference especially in terms of direction', the other one complained that 'the sound sometimes seemed to be coming from all directions and I was uncertain whether to turn left/right, only when I was close did the volume increase, but in a jump, not gradually'. These comments points out the drawbacks of our exaggeration process of 3D sound intensity in this AR environment and also to some extend support the results of Brungart (Brungart and Rabinowitz, 1999; Brungart, 1999). We designed the auditory environment so that when the user becomes closer to the virtual sound source in the three by three meters region, the 3D sound intensity is exaggerated. This will be maximum intensity when the user is at the position of the virtual sound sources (distance to the object is zero). In this case, the binaural cues play a very important role for distance and direction perceptions. Because of the closeness to the object when near zero, even a slight movement of the user may cause a totally opposite feeling of the source direction (from left to right). When the movement is very fast, the intensity and sound direction may change rapidly. This problem can be partly solved by designing a better sound-scape. However the situation when the user is overlaid on the sound source is still a difficult problem unless we apply some physical constraints on the user's sound near to virtual sound sources.

Chart (a) and (b) in Fig. 13 shows the subjects' responses to question four and question five, respectively. We can see from chart (a) that nine of 40 subjects claimed that they were 'always' be aware of co-players while five subjects claimed 'very often' and 18 claimed 'often'. This leads to the improvement to the feeling of collaboration. A total of 39 subjects agreed that 3D sound did increase their feeling of collaboration from the result of question two.

To examine the importance of 3D sound among all cues, subjects were required to choose the cues they thought to have helped them and to order them by the importance. 3D sound cues obtained the highest votes (36 votes) among all the cues. Move importantly, a total of 14 subjects regarded 3D sound cues as the most important cue while 'small talk' and 'vision cues' only obtained five and four votes respectively. This result suggests that 3D sound does play a very important role in improving the subject's feeling of collaboration. It could even dominate among all possible cues in AR environment, especially when other cues are restrained by hardware settings, for example, angle of view is restrained by HMD in common AR system. In this situation, the visual cues are restrained due to the narrow viewpoint of HMD, and hence their effects on collaboration are limited. On the contrary, the 3D sound can be generated from any points in the AR space. When 3D sound is applied to the user and given different meaning at different situations, it can play the most important role to help the user's feeling of 'human presence and collaboration'.

5. Conclusions

The purpose of this work is to investigate the effectiveness of 3D sound in AR environments We designed two experimental AR environments to study the effectiveness

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 024

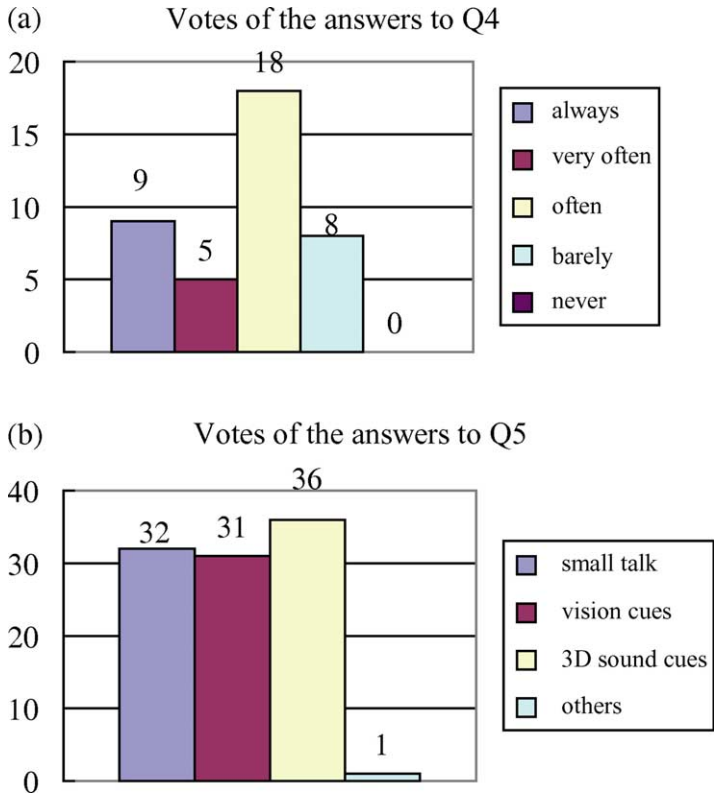

Fig. 13. Statistical charts of the answers to Question 4, 5 of Section 4.2.3. (a) The votes of the answers to question 4: Were you aware of your co-players in different places? We can see from the chart that in this AR environment, subjects can often be aware of co-player's position, however, this result is impossible if there was no 3D sound applied since the subject can only occasionally see co-player due to the narrow angle of view. (b) The votes of the answers to question 5: What has helped your feeling of 'human presence and collaboration'? It can be seen from this chart that 3D sound plays the most important role among all the cues. It gets even more votes than vision cues and voice communication cues.

both quantitatively and qualitatively. Two different tracking methods were applied to retrieve the 3D position of virtual sound sources. We examined the effects of 3D sound on improving depth perception and shortening task completion time. We also investigated 3D sound's effects on improving immersive and realistic perceptions, helping user to identify different spacial objects, and improving the subjective feeling of 'human presence and collaboration'.

The 3D sound generated by OpenAL (Loki Entertainment Software) applying HRTFs and simple modelling of environment is used in our experiments. Apart from its effects for the user on sound source localization, we tried to emphasize its effect on telling the depth (distance). Since intensity has been reported to be a key cue for distance perception, we exaggerated the intensity difference of sound sources in different depth by applying an additional scalar factor to the intensity of generated 3D sound. The user study result shows

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 025

that it is still 'intuitive'. These methods to implement 3D sound in relative small AR environments provide a necessary foundation for the incorporation of 3D sound with visual AR environments.

However, our studies beg many additional questions. For example, how to set the intensity scale so that it can maximize 3D sound's ability to tell depth and meanwhile preserve realistic perception for users? Moreover, as reported, for distance perception not only the intensity but also a much stronger cue, the ratio of the energies of direct and reflected sound is of importance (Middlebrooks and Green, 1991; Bronkhorst and Houtgast, 1999). How will this ratio affect the depth perception in AAR? If the cues of background and drastic head movements are allowed, to what extend will 3D sound be effective to help depth judgment, comparing to other visual cues? When the binaural localization cues for nearby sources (Brungart and Rabinowitz, 1999; Brungart, 1999) are taken into account, how to design the sound-scape to 'display' the 3D sound correctly and intuitively when the user is very close to or even right overlapped on the sound source?

Although many questions remain unanswered, our studies show that the addition of 3D sound can be a very effective way to complement the user experiences in AR environments. It helps depth perception, task performance and facilitates collaborations between users. Moreover, it enables a more realistic environment and more immersive feeling of being inside a both visual and auditory augmented reality world.

Acknowledgements

This work is supported by Defence Science and Technology Agency (DSTA), Land Systems Division, Singapore.

References

Avionique S, Multipurpose auditory display for 3-d hearing applications. Online. http://www.cordis.lu/esprit/src/22352.htm.

Azuma, R.T., 1997. A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6 (4), 355–385.

Barfield, W., Cohen, M., Rosenberg, C., 1997. Visual, auditory, and combined visual-auditory displays for enhanced situational awareness. The International Journal of Aviation Psychology 7 (2), 123–138.

Begault, D., 1994. 3-D Sound for Virtual Reality and Multimedia. AP Professional, Boston.

Bill G., Keith M. Hrtf measurements of a kemar dummy-head microphone, Tech. Rep. 280, MIT Media Laboratory Perceptual Computing Technical Report, May 1994.

Billinghurst, M., Bowskill, J., Jessop, M., Morphett, J., 1998. A wearable spatial conferencing space Proceedings of ISWC 98 1998, pp. 76–83.

Bronkhorst, A.W., Houtgast, T., 1999. Auditory distance perception in rooms. Nature 397, 517–520.

Brungart, D.S., 1999. Auditory localization of nearby sources. iii. stimulus effects. Journal of the Acoustical Society of America 106 (6), 3589–3602.

Brungart, D.S., Rabinowitz, W.M., 1999. Auditory localization of nearby sources. Head-related transfer functions. Journal of the Acoustical Society of America 106 (3), 1465–1479.

Cheok, A.D., Yang, X., Zhou, Z., Billinghurst, M., Kato, H., 2004. Touch-space: Mixed reality game space based on ubiquitous, tangible, and social computing. Personal and Ubiquitous Computing 2002. Vol. 6 2004 pp. 430.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 026

Cohen, M., Aoki, S., Koizumi, N., 1993. Augmented audio reality: telepresence/VR hybrid acoustic environments Proceedings of the Workshop on Robot and Human Communication 1993, pp. 3.61–64.

Eckel, G., Banissi, B., Khosrowshahi, F., Sarfraz, M., Ursyn, A., 2001. Immersive audio-augmented environments—the listen project. Proceedings of the Fifth International Conference on Information Visualization (IV2001) 2001.

Feiner, S., MacIntyre, B., Haupt, M., Solomon, E., 1993. Windows on the world: 2d windows for 3d augmented reality: 2d windows for 3d augmented reality Proceedings of UIST '93 1993, pp. 1.45–1.55.

Fuhrmann, A., Hesian, G., Faure, F., Gervautz, M., 1999. Occlusion in collaborative augmented environments. Computers and graphics 23, 809–819.

Gaver, W., 1986. Auditory icons: using sound in computer interfaces. Human Computer Interaction 2 (1), 166–177.

Gaver, W.W., Smith, R.B., O'Sheak, T., 1991. Effective sounds in complex systems: the arkola simulation. Proceedings of CHI '91 1991, 85–90.

InterSense, Is-900. online. http://www.intersense.com/.

Kanbara, M., Yokoya, N., 2002. Geometric and photometric registration for real-time augmented reality. IEEE and ACM International Symposium on Mixed and Augmented Reality 2002.

Kanbara, M., Fujii, H., Takemura, H., Yokoya, N., 2000. A stereo vision-based augmented reality system with an inertial sensor Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR) 2000, pp. 97–100.

Kato, H., Billinghurst, M., 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system Proceedings of the Second IEEE and ACM International Workshop on Augmented Reality 1999, pp. 85–94.

Langendijk, E.H.A., Bronkhorst, A.W., 2000. Fidelity of three-dimensional sound reproduction using a virtual auditory display. Journal of the Acoustical Society of America 107 (1), 528–537.

L. Firelight Technologies, Pty, fmod, music and sound effects system. [Online]. Available: http://www.fmod.org/.

Loki Entertainment Software, Openal SDK. [Online]. Available: http://www.openal.org.

Makino, H., Ishii, I., Nakashizuka, M., 1996. Development of navigation system for the blind using gps and mobile phone combination Proceedings of 18th Conference of IEEE, EMBS, vol. 18 1996, pp. 1–5.

Mereu S.W., Improving depth perception in 3d interfaces with sound, Tech. Rep. CS95-35, Waterloo, Canada, 1995.

Middlebrooks, J.C., Green, D.M., 1991. Sound localization by human listeners. Annual Review of Psychology 42, 135–159.

Milgram, P., Takemura, H., Utsumi, A., Kishino, F., 1994. Augmented reality: a class of displays on the reality-virtuality continuum. Telemanipulator and Telepresence 2351.

Mynatt, E.D., Back, M., Want, R., Frederick, R., 1997. Audio aura: light-weight audio augmented reality Proceedings of the UIST '97 User Interface Software and Technology Symposium 1997, pp. 21.1–2.

Neumann, U., You, S., 1999. Natural feature tracking for augmented reality. IEEE Transactions on Multimedia 1 (1), 53–64.

Poupyrev, I., Airola, C., Billinghurst, M., Kato, H., 2000. Augmented groove: collaborative jamming in augmented reality. ACM SIGGRAPH Abstracts and Applications 2000, 2000, 77.

Schmandt, C., 1998. Audio hallway: a virtual acoustic environment for browsing Proceedings of UIST '98 1998, pp. 1.63–70.

Stockburger D.W. Introductory statistics: Concepts, models and applications. online, 1998. http://www.psychstat.smsu.edu/introbook/sbk00.htm.

Sutherland, I.E., 1968. A head-mounted three dimensional display. Proceedings of the AFIPS Fall Joint Computer Conference 33, 7.57–7.64.

Takagi, A., Yamazaki, S., Saito, Y., Taniguchi, N., 2000. Development of a stereo video see-through hmd for ar systems Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR) 2000, pp. 68–77.

Wang, W., Yang, X., Cheok, A.D., 2002. Touch space: an embodied computing mixed reality game space. International Workshop on Entertainment Computing (IWEC 2002) 2002, 4.55–4.62.

Wenzel, E.M., 1994. Spatial sound and sonification, in: Kramer, G (Ed.), Auditory display: sonification, audification, and auditory interfaces, pp. 127–150.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 027

Wenzel, E.M., 1996. Research in virtual acoustic displays at nasa Proceedings of the Simulation Technology and Training Conference (SimTecT'96) 1996, pp. 85–90.

Wenzel, E.M., Wightman, F.L., Foster, S.H., virtual, A., 1988. A virtual display system for conveying three-dimensional acoustic information Proceedings of the Human Factors Society, 32 1988, pp. 86–90.

Wloka, M.M., Anderson, B.G., 1995. Resolving occlusion in augmented reality Proceedings of the 1995 symposium on Interactive 3D graphics 1995, pp. 5–12.

Zhou, Z., Fabiz, F., Chen, X., Cheok, A.D., 2002. Magic music desk: a multi-modal embodied interactive desk International Workshop on Entertainment Computing (IWEC2002) 2002, pp. 4.23–30.

Downloaded from https://academic.oup.com/iwc/article/16/5/989/761559 by New York University user on 03 April 2023

Meta Platforms, Inc.
Exhibit 1107
Page 028