BEHAVIOUR & INFORMATION TECHNOLOGY, NOVEMBER–DECEMBER 2004, VOL. 23, NO. 6, 435–446

Taylor & Francis
Taylor & Francis Group

Using 3D sound as a navigational aid in virtual environments

RYAN GUNTHER†, RICK KAZMAN‡ and CAROLYN MACGREGOR§

†Department of Computer Science; University of Waterloo, Canada; e-mail: rtg@sfu.ca

‡University of Hawaii and Carnegie Mellon University, USA

§Department of Systems Design Engineering, University of Waterloo, Canada

Abstract. As current virtual environments are less visually rich than real-world environments, careful consideration must be given to their design to ameliorate the lack of visual cues. One important design criterion in this respect is to make certain that adequate navigational cues are incorporated into complex virtual worlds. In this paper we show that adding 3D spatialized sound to a virtual environment can help people navigate through it. We conducted an experiment to determine if the incorporation of 3D sound (a) helps people find specific locations in the environment, and (b) influences the extent to which people acquire spatial knowledge about their environment. Our results show that the addition of 3D sound did reduce time taken to locate objects in a complex environment. However, the addition of sound did *not* increase the amount of spatial knowledge users were able to acquire. In fact, the addition of 3D auditory sound cues appears to *suppress* the development of overall spatial knowledge of the virtual environment.

1. Introduction

Careful consideration must be given to the design and usability of virtual environments, otherwise much of the potential for applications of virtual reality will be lost. One important design criteria is to make certain that adequate navigational cues are incorporated into complex virtual environments so that users can quickly orient themselves and can navigate throughout the environment with confidence and efficiency. The majority of the previous research on navigation in both real and virtual environments has focused on either the physical layout of the environment or the addition of visual navigational aids. There has been little work exploring the use of auditory cues as navigational aids. Auditory cues, as we will show, can be an important component of virtual environments because, due to the limitations of processing and display hardware, these environments are significantly less visually rich than real world environments.

Auditory information, when designed to complement the visual environment, is natural and people are innately comfortable with it – its use requires no training. There are several reasons why we feel auditory information is a good choice for augmenting navigational information. One advantage of auditory cues is that they provide a sense of spatial context. Even without head movement auditory information can be perceived in all directions, and we can track and selectively attend to multiple simultaneous audio streams (sometimes called the 'cocktail party' phenomenon). This is in contrast to visual information, where a human's natural field of view is restricted to about 150° horizontally. In addition, auditory cues complement visual cues (e.g. we can both see a clock and hear it ticking), both by providing information redundancy and by enhancing the user's sense of 'presence'. Finally, auditory cues are useful as navigational aids because they are complementary to the visual stream of information and do not require conscious choices of attention, in contrast to additional purely visual aids such as maps.

For these reasons, we hypothesized that the addition of 3D sounds to a visually sparse but structurally complex virtual environment would improve the ability of people to both find their way around the environment and to acquire *survey knowledge* (knowledge of object locations with respect to a global coordinate system). To test these hypotheses we constructed a virtual world and ran an experiment to compare users' navigational, or

Behaviour & Information Technology
ISSN 0144-929X print/ISSN 1362-3001 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01449290410001723364

wayfinding, performance (speed of locating a desired destination) in a sound enhanced environment with that of users in the same environment, but with no sounds.

In the experiment, we compared the sound enhanced and 'silent' conditions with respect to their effects on navigational performance. In each condition, users were asked to locate several objects and their time to find each object was recorded. We also evaluated the amount of survey knowledge users were able to acquire in each condition. We did this by having each user identify rooms on a map of the environment (by correctly identifying some of the contents of the room) after they had explored it for several minutes.

As expected, we found that the addition of 3D sounds to the virtual environment did improve a user's ability to navigate to specify locations in the virtual environment. However, somewhat surprising was that the presence of 3D sounds did *not* appear to enhance the acquisition of survey knowledge, and may even have *suppressed* its development. This has important implications when designing virtual environments for training, walk-throughs, and other applications, as we will discuss.

2. Background

There has been a substantial amount of research in the area of spatial knowledge theory. However, little of this research has been directly applied to the acquisition of spatial knowledge in virtual environments. We will briefly survey the existing literature in this area. A more complete survey of the literature can be found in (Gunther 1997).

2.1. *Spatial knowledge*

Goldin and Thorndyke (1982) have subdivided spatial knowledge into three categories: landmark knowledge, procedural knowledge, and survey knowledge. Each type of knowledge is focused on different attributes of an environment, is acquired through different functional experiences with the environment, and is useful for navigational different tasks.

2.1.1. *Landmark knowledge*: Landmark knowledge is information about the important visual details in an environment. Objects that will become landmarks are those that are visually distinctive or have personal meaning (Lynch 1960). Some attributes of an object that may distinguish it are architectural style, size, and colour. As well, any object that provides directional information is likely to become a prominent landmark (Satalich 1995).

2.1.2. *Procedural (route) knowledge*: Procedural or route knowledge has an egocentric (inside-out) frame of reference and represents information about the sequence of actions that are needed to follow specific routes. This information includes knowledge of critical points along a route where turns occur, and the action that is required at each point. Procedural knowledge implies knowing the approximate distance between route segments, the direction of turns, and the ordering of landmarks.

2.1.3. *Survey knowledge*: Survey knowledge has an exocentric (outside-in) viewpoint and represents object locations and inter-object distances with respect to a fixed, global coordinate system. Survey knowledge allows people to estimate distances between landmarks and to infer alternate routes that have never been travelled.

Successful *wayfinding* requires the use of all three types of spatial knowledge.

2.2. *Wayfinding*

Wayfinding is defined by Peponis *et al*. (1990) as 'how well people are able to find their way to a particular destination without delay or undue anxiety.' Landmarks are used by a navigator to acquire and maintain orientation, as well as to recognize destinations. Route knowledge is needed to follow a route, and survey knowledge is required to choose the most appropriate route.

The central hypothesis of the research presented here is that 3D spatialized sound not only aids a user in wayfinding in a virtual environment, but also should enhance the navigator's acquisition of route and survey knowledge over that which could be obtained via a silent virtual environment.

2.3. *3D (spatialized) sound*

Begault (1993) defines spatialized sound as a technique where:

> ...the outer ears are either directly implemented or modeled as digital filters. By filtering a digitized sound source with these filters, one can potentially place sounds anywhere in the virtual space about a headphone listener.

Adding 3D sound to a virtual environment has several advantages. It has been shown to increase a user's sense of presence (Hendrix and Barfield 1995); presence is the degree to which a person feels that they are actually *in* the environment. Sound can provide information that is redundant to that provided by visual cues. This makes it less likely that a person will misinterpret or lose important information. As well, a person's focus of attention for sound can be switched without noticeable physical effort, unlike vision that often requires eye or head movement. This is useful when many things need to be monitored or attended to at one time. 3D sound can also provide information about things that are happening in a 360° circle around a person.

It has only recently become convenient and economically feasible to generate 3D sound in real time on inexpensive hardware platforms. As a result, little work has been done exploring the use of 3D sound in practical applications. Now that 3D sound generation systems are less expensive they are becoming more widely available, and are being found effective for a variety of applications. For example, a study conducted by Begault (1993) found that when commercial airline crewmembers used a 3D auditory display to acquire spatial information about surrounding aircraft they were approximately 2.2 s faster in locating aircraft than crewmembers who used only one-earpiece headsets. Walker and Lindsay (2003) in a study of the effects of 'beacon sounds' in a navigational environment showed that these, when properly designed, can have a significant effect on performance. Tran *et al.* (2000) have studied the effects of beacon characteristics, both speech and non-speech, and have determined a set of auditory characteristics that are significantly correlated to localization performance. A number of studies for the visually impaired (Loomis *et al.* 1990, Mereu and Kazman 1996, Helal *et al.* 2001) have similarly shown that 3D sound, when properly designed, can have a positive effect on localization, but studies such as (Grohn *et al.* 2003) indicate that auditory navigation alone is less efficient than visual or a combination of the two.

A study performed by Darken and Sibert (1993) attempted to examine the usefulness of 3D sound as a navigational aid. In that study, a spatial audio cue was set to a steady positional tone and treated as an acoustic landmark. It was not audible throughout the entire environment, but when it became audible people appeared to use it for rough direction finding and it had the effect of enlarging the target area. The effectiveness of 3D sound as a navigational aid for enhancing survey knowledge was not the primary focus of Darken's study. Building on the findings of Darken, we decided to design an experiment to explore the role that 3D sound cues play in wayfinding and the acquisition of spatial knowledge of a virtual environment. Our main objective in this experiment was to determine if assigning 3D sounds to specific objects in a virtual environment decreases the amount of time it takes people to locate these objects (route knowledge). We also wanted to see if the addition of 3D sounds affects the number of objects from the environment a person is able to recall (landmark knowledge), and their ability to place those objects on a map of the environment (survey knowledge).

3. Method

A complete discussion of the experimental design and procedures can be found in (Gunther 1997). Here, we will only discuss the most important aspects of the design.

3.1. *Design of the virtual environment*

The emphasis of our study was on examining the effect that spatialized auditory cues might have on wayfinding and the acquisition of spatial knowledge of the test environment. As the focus was not on the direct comparison of visual vs. auditory cues, the test environment was intentionally designed to be visually sparse in order to better isolate the effects of the spatialized sounds on navigational performance.

To answer the research objectives we constructed a maze consisting of twelve rooms. Each room contained a set of distinct objects, and each room had a 3D sound associated with one of the room's objects. Collision detection was implemented such that participants were unable to walk through walls or room objects. Much like an office-cubicle environment, sound could be heard through the walls, and the volume of a sound source increased as the participant got closer to the associated object. None of the sounds could be heard throughout the entire environment.

Figure 1 shows a map of the virtual environment used in our study. The black arrow, situated in the lower middle of the environment, indicates the starting position and orientation of participants for each of the object search tasks. The hallways in the environment had grey floors, and the floor in each of the rooms was yellow. The ceiling was black everywhere, and all interior walls were beige. Each of the four exterior walls was a different colour: north was green, west was red, south was blue, and east was yellow. Coloured geometric shapes were placed on walls in the environment at most of the intersection points. These shapes were squares, diamonds, and circles and were one of three colours: red, green, or blue. There were a total of nine different geometric landmarks. The visual land-

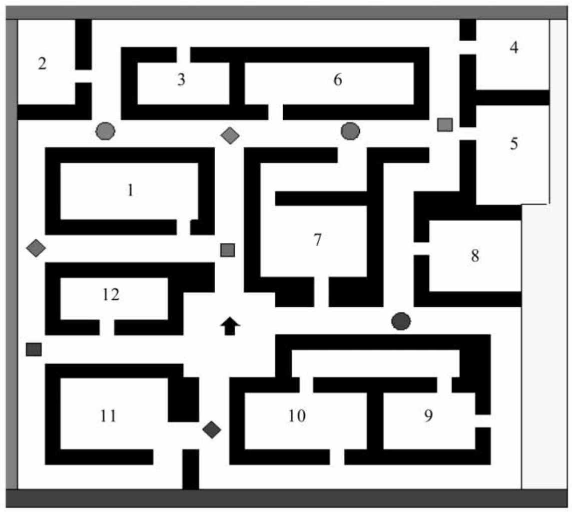

Figure 1. Experimental environment layout, showing the auditory objects in each room.

marks (i.e. coloured geometric shapes and walls) were included for several reasons:

- It would have been extremely unrealistic to have participants immersed in an environment that contained no visual cues that distinguished their location. Even the hallways of office buildings contain distinguishing features such as art on the wall, door numbers, plants, etc;
- also, if the environment contained no visual cues then it was assume that the navigational tasks would take longer, hence requiring the participants to be immersed in the virtual environment for a longer period of time. This would increase the potential for 'simulator sickness' (which a small number of subjects experienced even in the design that was eventually used).

The menagerie of stimulus objects was selected so as to be distinctive and memorable from either an auditory or visual perspective. Through a pilot study, all objects and sounds were subjectively determined to be recognizable and easily distinguished from each other. Exactly one 3D sound source was placed in each room, and inside each room there was an object, or set of objects, that was clearly associated with this sound (the *auditory objects*). As well, there was at least one object in each room that had *no* obvious association with any sound in the environment (the *non-auditory object*).

As we anticipated an auditory cuing advantage, we wanted to reduce the chances that those participants exposed to the auditory cues would be able to recall the non-auditory objects indirectly through association with the auditory objects. For each room, the objects were selected such that there were no clear associations between the auditory and non-auditory objects. At the same time, the size or content of the objects in the room was selected so as to foster visual mnemonics for all participants.

The rooms, their visible objects, and their associated sounds are listed in Table 1.

The placements of the objects were such that the participant had to enter the rooms and look about in order to locate all of the test objects. For example, in Room no. 12 in the 'sound on' condition, the participant would be able to hear the ticking of a clock from the hallway. However, he would have to move into the room in order to see the pictures of the clock, the teapot on the coffee table and the two couches.

3.2. *Stimuli and apparatus*

Two PCs created the virtual environment used in this study. One computer was used to generate the graphics and to control movement through the virtual environment. The Sense8 3D authoring package, World Up, was used to construct the virtual environment, and the World Up Player was used to allow people to navigate throughout the environment. The second PC was used solely to produce the 3D sounds. The QMixer95 real time 3D audio mixer was used to generate the 3D sounds. The two PCs were connected via a network.

A head mounted display (HMD), the Virtual i-o glasses, was used to display the virtual environment to participants during the experiment. The HMD had a 30° field of view, a resolution of 180 000 pixels per LCD panel, and a pair of stereo headphones, through which the 3D sounds from the environment could be heard. The head-mounted display was also equipped with head tracking, so that a left or right head rotation caused a change in the view of the virtual environment. We disabled the pitch and roll headtracking capabilities of the HMD because during the experiment there was no reason for participants to look up or down, and we felt that limiting the amount of head motion would prevent participants from developing simulator sickness. A 2-button mouse was also used to allow participants to move through the virtual environment.

Movement during the experiment was constrained to the horizontal dimension; no movement in the vertical direction was possible. Pushing the left mouse button had the same effect as a left head rotation, and pushing the right mouse button was equivalent to a right head rotation. Simultaneously holding down both buttons allowed a participant to move forward; no sideways or backwards motion was possible. Speed of movement

Table 1. Objects in the environment (Auditory objects are in italics).

Room	Object Set	Room Sound
1	*Violin*, Hammer, Bar table, Bar stools	*Violin playing*
2	*Telephone*, Desk, Bookshelf, Sunglasses	*Telephone ringing*
3	*Helicopter*, Baseball bats	*Helicopter rotary blades turning*
4	*Bowling alley*, Garbage cans	*Bowling pins falling over*
5	*Birds*, Balloon, Lifeguard chair	*Seagulls calling*
6	*Ping-pong table*, Dart board	*Ping-pong ball batted & bouncing*
7	*Horses*, Picnic tables; Mailbox	*Horse whinnying*
8	*Keyboard*, Chairs, Wrapped gift	*Scale played on piano*
9	*Frogs*, Plants	*Frogs croaking*
10	*Children's play gym* (Swings, Monkey bars, See-saw), Bench	*Children laughing*
11	*Dogs*, Cage, Barbell	*Dogs barking*
12	*Clocks*, Coffee table, Couches; Teapot	*Clock ticking*

was fixed for all participants, representing a moderate walking speed. A small sphere was used to indicate the current direction of forward motion and was always visible in the middle of the screen. The sphere (which was coloured blue in the environment), is located in the centre of the screen image, as can be seen in figure 2.

Figure 2. Sample screen shot of the environment (Room 12).

3.3. *Participants*

The participants in the experiment were primarily graduate and undergraduate students from the University of Waterloo, ranging from 16–35 years of age. A total of 50 people were recruited for the experiment, 36 men and 14 women. Participants were randomly assigned to one of three conditions: No Sound (for both blocks of trials), Full Sound (for both blocks of trials), Partial Sound (sound on for first block of trials, sound off for second block of trials). As males tend to be faster at navigating through virtual environments (Keppel 1991, Galea and Kimura 1993) the number of female participants was balanced across the three conditions.

3.4. *Procedure*

The experiment took place in a quiet room situated in the Computer Graphics Lab, at the University of Waterloo. During navigation of the virtual environment the lights in the room were turned off and participants were seated, wearing the head mounted display. The experiment took approximately $1^1/_2$ h and each participant was asked to perform several tasks during the course of the experiment. As well, participants were asked to verbalize their thoughts throughout the various stages of the experiment and these were recorded on audiocassette. The purpose of asking participants to verbalize their thinking was to provide an indication that those participants exposed to the auditory cues were aware of them. It also allowed the researcher to monitor whether the participant was making progress in the environment or was hopelessly lost – something that could not be derived from performance time alone.

First, each participant was asked to fill out a background questionnaire so that we would know if our experimental conditions were balanced according to participant demographics. Then all participants were given a training session to become familiar with the equipment. Each participant was asked to explore a practice environment consisting of five rooms until they felt comfortable using the HMD and mouse to navigate throughout the environment. None of the participants needed more than 5 min before they felt competent using the HMD and mouse to navigate through the environment. The practice rooms had the same coloured walls, floors, and ceilings of the test environment but did not contain any of the objects included in the main environment.

After the training session, each participant was shown a coloured map of the environment for 2 min (a coloured version of figure 1). The coloured visual cues (coloured exterior walls and geometric shapes) were included on the map, and the map was oriented north up. The labels for the rooms and objects were not included on the map. The participant then donned the HMD and was teleported (moved directly) to each of the 12 rooms of the virtual environment. All participants visited the rooms in the same order and had 14 s to look around in the room. This was to ensure that all participants had been exposed to all rooms and objects prior to the start of the timed trials. They were then given 8 min to navigate throughout the environment on their own. There were two reasons why this phase was included in the experiment. First, to determine if the addition of spatialized sound to a virtual environment has a significant effect on peoples' exploration strategies or patterns of movement. Second, to provide participants with some time to become familiar with the layout of the environment. After the free exploration phase, participants were asked to perform eight timed trials (two blocks of four trials each). For each trial, the participant was to locate the target object within the object as quickly as possible. All trials started at the designated starting point. The participant initiated the trial by pressing the space bar on the keyboard to display an image of the target object that the participant was to find for that trial. Participants indicated they had located the object by colliding into it in the virtual world. A message indicating success was displayed on the screen after each object was found. The participant was then teleported to the starting point for the next trial. A maximum duration of 4 min was set for each of the timed trials. This was done to avoid participants wandering indefinitely in search of an object, and to minimize exposure time so as to reduce chances of motion sickness. If the target object was not located in 4 min, then a message was displayed indicating that the trial time was up and the participant was returned to the starting position for the next trial. Participants had a short break between the two blocks of trials.

We recorded both their routes and the time they took to find each object. The test environment was divided into sectors so that we could track how often the participant visited various parts of the environment. In addition to recording routes and times, participants were asked to verbalize aloud their decision making while moving through the environment. This allowed the researcher to be aware of whether the participant was spending time in a particular sector because they were still looking about or whether they were lost.

The target objects were selected such that eight of the 12 rooms served as destinations. Four of the rooms were purposely not used as destinations. This was to allow us to check to see whether participants would visit and recall information from non-targeted rooms. From the participant's perspective the eight target objects appeared to be selected at random. In fact, all participants were shown the same four targets and in the same order for each block of trials. Targets were selected so as to represent comparable destination sectors within the world.

In Group A targets were a telephone (Room 2), a horse (Room 7), a ping-pong table (Room 6) and a frog (Room 9). In Group B targets consisted of a dog (Room 11), a helicopter (Room 3), a swing set (Room 10), and a bowling alley (Room 4). Rooms 1, 5, 8, and 12 were not specifically targeted.

To ensure that results were not driven due to one block of targets being inherently easier to locate than another, order of groups of targets was counterbalanced for the No Sound and Partial Sound conditions. The counterbalancing was restricted to those groups who experienced the world without sound. In other words, half of those participants received the Group A targets for their first block of trials, and the other half received Group B targets for their first block of trials.

After completing the timed trials, each participant was asked to perform a pen-and-paper task: to recall as many objects from the environment as possible and place each object in the correct room on a coloured map of the environment oriented with north up, similar to the one they were shown at the beginning of the test session. Finally, participants were asked to write the Guilford-Zimmerman test of spatial orientation (Guilford and Zimmerman 1948) and complete a short questionnaire that asked for their impressions of the experiment. We administered the spatial orientation test after the timed trials as we did not want participants to vary their behaviour based on their self-assessment of their performance on the Guilford-Zimmerman test.

3.5. *Conditions*

As mentioned above, participants were randomly assigned to one of three experimental conditions called: *NoSound, FullSound*, and *PartialSound*. The difference among the three conditions was the amount of time that 3D sounds were heard during the experiment. Participants placed in the NoSound condition were only presented with the visual environment and never heard 3D sounds. Participants in the FullSound condition were presented with both the visual and auditory environments and were able to hear sounds during the entire experiment. In the PartialSound condition participants were able to hear sounds for the entire exploration phase, and for the first block of four trials. The sound cues were turned off for the second block of trials.

4. Results

All data analyses were carried out using SPSS. Unless otherwise stated, significant omnibus F's were further evaluated using LSD *post-hoc* comparisons. As this is exploratory research, in that few others have investigated the effects of 3D audio cues as navigational aids in virtual environments, and given the relatively small sample sizes, there is justification in using the LSD over the more stringent Tukey or Sheffé tests (Keppel 1991).

The data for five participants were not included in the analyses. Three of the participants experienced moderate symptoms of motion sickness (e.g. complaints of dizziness and headaches), and two others were unable to complete the required tasks, one due to fatigue unrelated to the study and the other due to a lack of understanding of the task requirements. The data for the remaining 45 participants were analysed by auditory condition: NoSound (13 males, 4 females), FullSound (6 males, 4 females), and PartialSound (13 males, 5 females). The three participants who experienced moderate symptoms of motion sickness were all from the FullSound condition. We have no reason to suspect that there was any correlation between this condition and the motion sickness, although we cannot prove this conclusively.

4.1. *Group demographics*

A series of one-way ANOVAs was carried out on the following participant demographics: self-assessment of navigation ability, average hours/week playing Doom-like games, average hours/week listening to music, average hours/week playing a musical instrument, average hours/week using a computer, and general familiarity with virtual reality. The groups did not differ significantly from one another on any of the aforementioned self-reported measures ($p < 0.05$). Furthermore, while the males in the study generally scored higher than the females on the Guilford-Zimmerman test of spatial orientation ability, there were no significant differences between the three experimental groups on this measure. Thus, the groups appeared to be reasonably equivalent prior to their exposure to the virtual environment.

4.2. *Comparing blocks of objects for ease of location*

As described previously, each participant was required to find eight objects from the virtual environment as quickly as possible, and their times to locate these objects were recorded. For each participant, the average mean time to locate each of four objects within each of the two blocks of trials was calculated. The objects that had been assigned to Block 1 and 2 were balanced in terms of their placement within the VE (2 interior rooms and 2 exterior rooms). Block means were analysed rather than means for individual object searches, as we were more interested in average time to complete trials than in the particular route strategies (as related to search time) used for locating individual objects.

In order to test that one set of objects was not inherently faster to locate based on visual information than the other set, the order of the blocks of objects was balanced for the NoSound group. In addition, to test that one set of objects was not inherently faster to locate based on auditory cues the order of the blocks of objects was balanced for the PartialSound group. None of these planned comparisons proved to be significant ($p > 0.05$). Thus, any significant differences between the groups are deemed to be due to the overall experimental conditions and not due to the inherent characteristics of the objects themselves.

4.3. *Exploration strategy by auditory condition*

A series of one-way ANOVAs was carried out to ensure that the three auditory conditions were comparable in terms of their coverage of the virtual environment. For instance, we wanted to know whether the FullSound and NoSound groups differed in terms of the amount of time that they spent in the hallways and in the rooms. One might expect the FullSound group to spend more time in the hallways listening for the appropriate auditory cue related to the target object, while the NoSound group might be more inclined to venture into the rooms in search of the target. The sectors of the test environment were divided into Interior Rooms, Exterior Rooms, Interior Hallways, Exterior Hallways, and Start Location. None of these analyses were found to be significant ($p > 0.05$), suggesting that the groups were comparable in terms of their coverage of the test environment.

4.4. *Object search time by auditory condition*

To analyse object search time, a 3×2 ANOVA (Auditory Condition by Block of Trials) was run with blocks of trials representing a repeated measure. A significant two-way interaction was found [$F(2,42) = 4.981$, $p < 0.012$]. Figure 3 shows a plot of the means for this interaction. For Block 1, the faster groups were the FullSound (X = 63.9 s, SD = 23.02) and the PartialSound (X = 69.3, SD = 30.09). *Post hoc* comparisons revealed that the NoSound group (X = 90.2 s, SD = 37.08) took significantly longer to locate objects in the first block than did the FullSound group [LSD(25) = 26.29, $p < 0.05$]. This effect is further exemplified if the FullSound and PartialSound are treated as one group – as both of these groups received the 3D sound cues during the first block of trials while the NoSound group did not [Planned comparison: t(43) = 2.367, $p < 0.02$].

For Block 2, the average object search time for the NoSound group (X = 68.2 s, SD = 32.42) had improved such that it was no longer significantly different from the FullSound group (X = 62.6 s, SD = 17.6). Interestingly, the average object search time for the PartialSound group (X = 118.3 s, SD = 89.34) degraded such that it was now significantly longer than those for the NoSound group [LSD(33) = − 50.07, $p < 0.02$] and the FullSound group [LSD(26) = 55.60, $p < 0.03$].

4.5. *Object recall task*

After each participant completed the object search trials they were asked to recall as many objects as possible from the environment and place each object in the correct location on a map of the environment. There were a total of 12 Auditory objects and 22 NonAuditory objects. Figure Y shows the mean number of Auditory and NonAuditory objects recalled by participants in each of the three conditions.

As shown in figure 4, the NoSound group recalled comparable numbers of NonAuditory (X = 9.2, SD = 3.00) and Auditory objects (X = 9.8, SD = 1.01). The PartialSound group also recalled comparable numbers of NonAuditory (X = 9.5, SD = 4.08) and Auditory objects (X = 10.7, SD = 1.19). However, the FullSound group recalled significantly more Auditory objects (X = 11.0, SD = 0.82) than NonAuditory objects (X = 7.3, SD = 5.52), [t (9) = 2.31, $p < 0.05$].

Two planned one-way ANOVAs were conducted for each of the object types by auditory condition. The one-way ANOVA for Auditory objects recalled proved to be significant [$F(2,44) = 4.75$, $p < 0.02$]. *Post-hoc* comparisons found the NoSound group (X = 9.8, SD = 1.05) recalled significantly fewer Auditory objects than either the PartialSound group (X = 10.7, SD = 1.19, $p < 0.03$) or the FullSound group (X = 11.0, SD = 0.82, $p < 0.02$).

The one-way ANOVA for NonAuditory objects was not significant. This may be due to the large variance associated with the FullSound group on this measure [NoSound: X = 9.4, SD = 2.99; PartialSound: X = 9.5, SD = 4.08; FullSound: X = 7.3, SD = 5.52). Although the comparisons were not significantly different, it is interesting that those participants receiving the auditory cues for all of the trials recalled fewer of the NonAuditory objects than did those participants who had to spend at least some of their trials in a silent world.

4.6. *Object placement task (room identification)*

As well as recalling objects, participants were also asked to place as many objects as possible in the correct rooms on a map of the environment. The data examined for this analysis was the number of rooms each participant was able to correctly identify. There were a total of 12 rooms in the environment. A room was considered to be correctly identified if at least one of the objects from a room was placed in the room on the map of the environment. Participants rarely attempted to place an object in more than one room, and were more inclined to not place an object at all if they were not

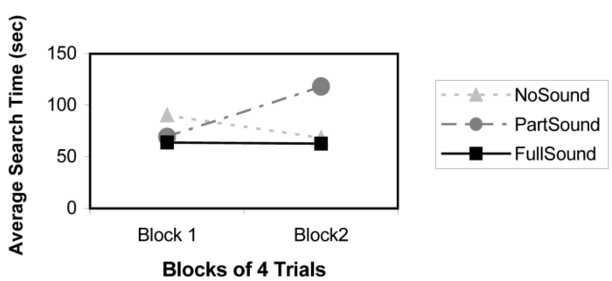

Figure 3. Object search time by auditory condition.

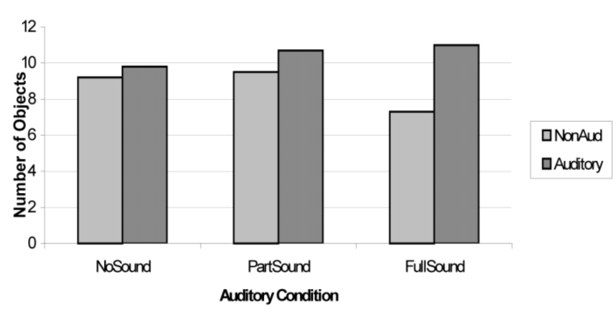

Figure 4. Object recall by object type.

certain of its location. As a result error rates were not used as a measure of performance due to low frequencies. In order to restrict object placement to the user's immediate recall of the environment, participants were asked to locate on the map only those objects that they had recalled. This was done to avoid situations wherein a participant might be able to locate initially unrecalled objects based on association with a recalled object.

The NoSound group (X = 7.1, SD = 3.07) correctly identified the most rooms on the map of the environment, followed by the PartialSound group (X = 5.2, SD = 4.29), and the FullSound group (X = 3.5, SD = 3.34). A one-way ANOVA by condition was found to be marginally significant [$F(2,44) = 3.08$, $p < 0.058$]. *Post-hoc* comparisons revealed that the NoSound group had correctly identified significantly more than the FullSound group [LSD (25) = 3.56, $p < 0.02$].

Further analysis was carried out to see whether the groups had differed in terms of their ability to identify target rooms and non-target rooms. As was found for the overall identification, the group identifying the most target rooms was the NoSound group (X = 5.5, SD = 2.28), followed by the PartialSound group (X = 3.9, SD = 3.08), and the FullSound group (X = 2.8, SD = 2.62). A similar pattern of results was found for the identification of the nontarget rooms (NoSound: X = 1.7, SD = 1.25; PartialSound: X = 1.5, SD = 1.46; FullSound: X = 0.7, SD = 0.82). While the one-way ANOVA for the nontarget rooms was not found to be significant, the one-way ANOVA for the target rooms was significant [$F(2,43) = 3.25$, $p < 0.05$]. *Post-hoc* comparisons verified that the NoSound group had identified more target rooms than did the FullSound group – even though the FullSound group had recalled significantly more Auditory objects than had the NoSound group [LSD (25) = 2.70, $p < 0.02$]. Figure 5 presents the overall results for room identification by

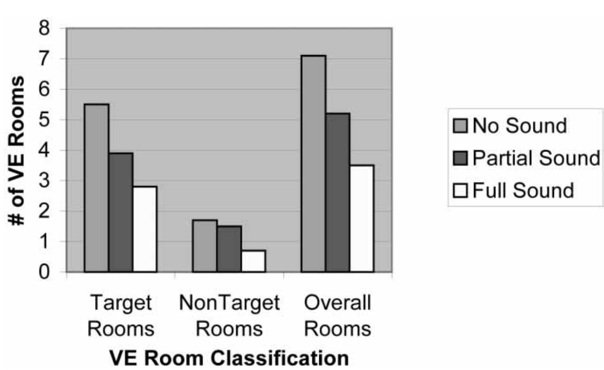

Figure 5. Room identification by auditory condition.

auditory condition.

We recognized that it is possible to remember a grouping or association of objects from the same room without being able to place that grouping in the exact location (room) on the map. For this reason, the number of average number of correct groupings was calculated by Auditory condition [NoSound: X = 5.7, SD = 0.71; PartialSound: X = 5.3, SD = 3.83; FullSound: X = 4.6, SD = 4.35]. No significant differences were found between the three auditory conditions in terms of grouping objects from the same room on the map.

5. Discussion

At the outset, we speculated that the richer 3D sound-enhanced environment would facilitate not only the acquisition of landmark knowledge in terms of objects recalled, but should facilitate the acquisition of route knowledge by way of faster target search time and the acquisition of survey knowledge by way of a more accurate mental map of the location of the objects within the virtual environment.

Comparisons of the experimental groups in terms of demographics suggested that the groups were comparable prior to their exposure to the virtual environment. Results of the comparisons of object order suggested that any improvement in ability to locate objects in the second half of the study was due to the experimental condition itself rather than the ease of target location of one set of objects over another. Bearing in mind that all participants had equal exposure to the environment prior to the experimental trials (in that they all received a tour of each of the twelve rooms and were shown a coloured map of the environment indicating the general layout in terms of outer and inner walls making up the hallways and 12 rooms), it is reasonable to assume that efficiency in search strategies during the experimental trials was not due to one group having spent more time in the environment than any other group.

We will discuss the findings in terms of route knowledge, landmark knowledge and survey knowledge as they correspond to the order of the measures taken: object search time, object recall, and object placement.

At the beginning of the experimental session, the participants had been exposed to the environment as a maze of 12 rooms. Therefore if they associated a distinct sound with a room and the objects in the room with that sound tag, then as long as they remembered which objects were located together, the participants in the sound conditions should be able to find a more efficient route to the room. In theory these participant should be able to shorten their routes, as they would not have to look into each room in order to locate the target object.

5.1. *Assessing route knowledge through object search*

When we examine the results from the first block of trials for the object search task, it was clear that the two groups who had received the 3D sound cues were able to locate the objects faster than the NoSound group ($p < 0.05$). If we had ended the experiment there, we might have concluded that the addition of 3D sound does in fact facilitate object search within virtual environments. Arguably the more interesting findings are derived from the analysis of the second block of trials. For Block 2, the NoSound group had shortened their search times such that they were able to locate the objects as quickly as the FullSound group. As the NoSound group did not have the benefit of the 3D sound cues, one must surmise that they were relying on visual cues to navigate within the virtual environment, and were developing more efficient strategies as they learned about the environment. Of note is the fact that the FullSound group did not show any improvement over time. It may be that they were already as efficient as they could be with the help of the 3D sound cues.

We are given a better glimpse of the strategies used by the FullSound group when we examine the results for the PartialSound group. When the PartialSound group had the benefit of sound during the first block of trials, they were as efficient in locating the target objects as the FullSound group. However, the PartialSound group took significantly longer to find the objects in the second block than did either the FullSound or NoSound groups ($p < 0.05$). Comments made by members of the PartialSound group suggest that they had been using the 3D sound cues as *auditory beacons* or landmarks for the various objects and had followed the sounds until they 'homed in' on the target object. Once these auditory beacons (landmarks) were turned off (Block 2) many commented that they were 'lost' or 'didn't know where they were going' without the sounds. This suggests that the addition of the auditory cues may have encouraged the participants to rely heavily on the sounds as homing devices and consequently they paid less attention to the visual cues in the environment than they would have if the sounds had not been present.

With only the object search time, we could not be sure that the PartialSound and FullSound groups had made use of the sounds in the same way to gain survey knowledge of the environment. It was possible that turning-off the sounds was a disruptive event and may have interfered with the development of a mental map of the environment for the PartialSound group. For that to be the case, we would have expected those participants to have developed inferior landmark and survey knowledge when compared to the participants who experienced the same sound (or non-sound) environment for all of the timed trials.

5.2. *Assessing landmark knowledge through object recall*

Recalling an object indicates whether the participant can remember items that make up that environment (both landmarks and details). Keep in mind that none of the participants had been told in advance that they would be asked to recall the objects that they had seen or to locate those objects on a map of the virtual environment. The fact that the NoSound group recalled almost an identical number of Auditory and NonAuditory objects is not surprising. Without the associated sounds, all of the objects in the environment would have been of comparable 'value' in terms of visual cues and associations. For example under the NoSound condition, the ping-pong table would have been of equal visual interest as the dartboard in the Games Room. However, under the FullSound and PartialSound conditions, the sound of a ping-pong ball being batted back and forth may have made the ping-pong table more memorable than the dartboard. In fact, the FullSound group was found to have recalled significantly more Auditory objects than NonAuditory objects ($p < 0.05$). Furthermore, the NoSound group recalled significantly less Auditory objects than either the FullSound ($p < 0.02$) and PartialSound ($p < 0.03$) groups. It is interesting that the PartialSound group recalled almost as many Auditory objects as the FullSound group, and almost as many NonAuditory objects as the NoSound group. It is as if having the benefit of the sound for the first four trials made those 12 objects memorable, while having the sound shut off for the last four trials forced the group to really look at what was in the room instead of just *hearing* what was in the room.

5.3. *Assessing survey knowledge through object placement*

Correctly locating those objects on a map of the environment requires that the user is aware of the spatial relationships between the various objects (survey knowledge). The first level of survey knowledge is the ability to group objects together by proximity – in this case a participant should be able to group objects that belong in the same room. The second level of survey knowledge is the ability to place individual objects or groupings into the right Cartesian location – in this case in the correct room. In terms of grouping of objects, there were no significant differences between the auditory conditions. However, in terms of correctly identifying rooms on the map through object placement, the NoSound group identified significantly more rooms than the FullSound group ($p < 0.02$). This was largely driven by the fact that the NoSound group was better able to identify the target rooms than was the FullSound group, despite the fact that the FullSound group had recalled more Auditory objects than the NoSound group ($p < 0.008$).

While not a significant improvement, the ParitalSound group also identified more rooms than the FullSound group. We conjecture that the PartialSound group performed better than the FullSound group on object placement because they were forced to pay more attention to the visual cues during their second block of trials. It is possible that the ParitalSound group benefited from the longer search times in that the inefficiencies may have forced them to examine the rooms more systematically before finding the target object. Consequently, they may have learned the contents and placement of the rooms better than the FullSound group.

So what explains the results that we have reported here? Previous research by Darken and Sibert (1993) found that the addition of a 3D sound to a virtual environment had the effect of enlarging the target area for the object with which the sound was associated. It appears that a similar effect occurred in this experiment and as a consequence, search directed primarily by sound cues became an efficient strategy and dramatically lessened the need to attend to visual cues. In effect, the 3D sound cues encouraged an efficient opportunistic navigation strategy of the environment. If an individual only has to listen for the correct cue and home in on it, then they don't need to thoroughly explore or even attend to the details of their environment. Such conditions allow individuals to get where they are going with very little conscious effort required and very little learning along the way. The NoSound group, on the other hand, benefited from being forced to do a more systematic exploration of their environment. While their initial trials took longer, they appear to have served to reinforce the location of objects within the environment.

In other words, it may have been the case that during the object search trials when participants who were receiving audio cues came within range of the enlarged target area their cognitive load was reduced to the point where the acquisition of survey knowledge was inhibited because it seemed irrelevant to them. If this was the case then the addition of 3D sounds to a virtual environment has the same affect that Moeser (1998) observed when signs were presented in a real world environment – the acquisition of wayfinding knowledge is actually inhibited.

What we were unable to answer from this experiment is whether the NoSound group would have continued to improve beyond that of the FullSound if we had added a third Block of trials. In order to keep the target search locations unique (i.e. one object per room) we were limited to working with less than 12 object searches – otherwise the participant could anticipate the location of the final object through process of elimination. Future studies should look at the effectiveness of 3D auditory cues as a function of the size or complexity of the virtual environment.

6. Conclusions

This study has focused on understanding the effects of 3D sounds on navigation in a virtual environment. The fact that the participants in each of the three Auditory conditions (NoSound, PartialSound, and FullSound) did not differ significantly in terms of spatial ability, or training, or overall exposure to the visual aspects of the virtual environment reinforces that our findings can be attributed to the role the 3D auditory sounds played in wayfinding and acquisition of spatial knowledge of the virtual environment.

The results of this study suggest that the addition of 3D sounds to a virtual environment can help a person locate objects and rooms faster than without 3D sound cues. However, the results also show that 3D sounds do not aid the acquisition of survey knowledge, and may even suppress its development.

Based on our findings, we recommend that designers of virtual environments carefully consider the combination of levels of navigational knowledge that can be gained or hampered by introducing additional cues within an environment. If the task goal is for the user to locate objects within a virtual environment as quickly as possible then including 3D sound cues may be of benefit. However, if the task goal is for user to gain a survey level of knowledge of the environment, then the inclusion of 3D auditory cues needs to be carefully considered. Further research might investigate whether a greater benefit could be derived from 3D sound cues if they are introduced after the user had a chance to learn the environment without them.

Acknowledgments

Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada, the University of Waterloo Institute for Computer Research, and the Computer Graphics Lab.

References

Begault, D. R. 1993, Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors*, **35**(4), 707–717.

Darken, R. P. and Sibert, J. L. 1993, A toolset for navigation in virtual environments. *UIST 93 Conference Proceedings* (ACM Press), pp. 157–165.

Galea, L. A. M. and Kimura, D. 1993, Sex differences in route-learning. *Personality and Individual Differences*, **14,** 53–65.

Goldin, S. E. and Thorndyke, P. W. 1982, Simulating varigation for spatial knowledge asquisition. *Human Factors*, **24**(4), 457–471.

Grohn, M., Lokki, T. and Takala, T. 2003, Comparison of auditory, visual, and audio-visual navigation in a 3D space. International Conference on Auditory Display, 6–9 July, 2003.

Guilford, J. P. and Zimmerman, W. S. 1948, The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology*, **32,** 24–34.

Gunther, R. 1997, The use of 3D sound as a navigational aid in virtual environments. *University of Waterloo Technical Report cs-97-30*.

Helal, A., Moore, S. and Ramachandran, B. 2001, Drishti: An integrated navigation system for visually impaired and disabled. Fifth International Symposium on Wearable Computers.

Hendrix, C. and Barfield, W., 1995, Presence in virtual environments as a function of visual and auditory cues. *IEEE Virtual Reality International Symposium*, 74–82.

Keppel, G. 1991, *Design and Analysis: A Researcher's Handbook* (Englewood Cliffs, NJ: Prentice Hall).

Lawton, C. A., Charleston, S. I. and Zieles, A. S. 1996, Individual and gender related differences in indoor wayfinding. *Environment and Behavior*, **28**(2), 204–219.

Loomis, J. M., Hebert, C. and Cicinelli, J. G. 1990, Active localization of virtual sounds. *Journal of the Acoustical Society of America*, **88,** 1757–1764.

Lynch, K. 1960, *The Image of the City*, (Cambridge: MIT Press).

Mereu, S. and Kazman, R. 1996, Audio enhanced 3D interfaces for visually impaired users, Proceedings of CHI '96, April 1996, 72–78.

Moeser, S. D. 1988, Cognitive mapping in a complex building. *Environment and Behavior*, **20,** 21–49.

Peponis, J., Zimring, C. and Choi, Y. K. 1990, Finding the building in wayfinding. *Environment and Behavior*, **22**(5), 555–590.

Satalich, G. A. 1995, Navigation and wayfinding in virtual reality: Finding the proper tools and cues to enhance navigational awareness. University of Washington Human Interface Technology Lab Technical Report TH-95-4.

Tran, T. V., Letowski, T. and Abouchacra, K. S. 2000, Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics*, **43**(6), 807–827.

Walker, B. N. and Lindsay, J. 2003, Effect of beacon sounds on navigation performance in a virtual reality environment. International Conference on Auditory Display, 6–9 July, 2003.