Impact of Binaural 3D Sound on Navigation Within a Virtual Environment

Nonnarit O-larnnithipong, Armando Barreto, Fatemeh Abyarjoo

Electrical and Computer Engineering Department. Florida International University. Miami, FL. USA
{nolar002, barretoa, fabya001}@fiu.edu

***Abstract*— This paper describes the execution and analysis of an experiment carried out to assess the impact of providing binaural 3D sound guidance during a search task within a virtual reality environment. The experiment compared guidance by full binaural 3D sound to guidance provided by a sound whose intensity (equal in both ears) decreases as the straight-line distance from the sound source increases. The results of the experiments with 8 human subjects reveal that providing full binaural 3D sound guidance resulted in a statistically significant reduction in the time required to complete the search task.**

***Keywords*—Virtual Reality, Binaural Sound, 3D Sound, Oculus Rift, Maze, Search, Navigation.**

I. INTRODUCTION

Virtual Reality (VR) applications continue to expand in areas of science, industry and entertainment. As more affordable and more powerful computational platforms continue to become available for the implementation of VR, more features are added to the VR environments, attempting to achieve the highest possible level of "presence" perceived by the user. That is, many contemporary VR implementations strive to provide an experience that is almost indistinguishable from "the real world" to the user. Beyond the subjective feeling of actual "presence" in the VR environment, modern VR systems aim at providing the user with the ability to perform tasks in the environment with as much proficiency as he or she would perform in the real world. In this context, it has become important to identify which are the features of a VR environment that are most critical in providing the perception of presence to a user and in allowing him / her to perform similarly as he / she would in the real world. For example, Dinh et al. [1] explored the potential benefits of providing multi-sensory cues (which included tactile, olfactory, audio and visual cues) to the user of a VR environment. Their results "strongly indicate that increasing the modalities of sensory input in a virtual environment can increase both the sense of presence and memory for objects in the environment." Interestingly, they also found that "increasing the level of visual detail did not result in an increase in the user's sense of presence or memory of the environment." Gröhn et al. [2] compared 3 types of guidance provided to users of a VR environment in a search task: Auditory, Visual and Audiovisual navigation. They found that "Audiovisual navigation was clearly the most efficient. Visual Navigation was second and the auditory navigation the least efficient."

With respect specifically to the sound delivery methods used in VR environments, there has been continued interest in determining what might be the impact of providing highly realistic spatialized sound. Loomis et al. [3] studied 3 different acoustic types of delivery of orientation information to blind individuals navigating in a real environment and found that the "virtual display mode [binaural spatialized audio] fared best in terms of both guidance performance and user preferences." Lokki et al. [4] executed a similar navigation experiment, but within a VR environment and found that audio cues can be effectively used to navigate a virtual environment and that even "simple models of spatial hearing give enough cues for auditory navigation." Zhou et al. [5] extended this type of study to an augmented reality (AR) environment and concluded that "the addition of 3D sound can be a very effective way to complement the user experiences in AR environments."

Binaural hearing is known to be a major asset for human sound source localization, and therefore, for the completion of navigational tasks guided by auditory beacons. In fact, systematic studies have confirmed that important degradation of the human ability to localize the source of a sound occurs when binaural hearing is disrupted. Kobler and Rosenhall [6] studied the localization ability of 19 subjects with mild-to-moderate hearing loss in order to evaluate the advantages and disadvantages of bilateral and unilateral hearing aid (HA) fittings. They found that "…bilateral HAs preserved the subjects' horizontal localization, whereas unilateral amplification decreased their horizontal localization abilities." They also concluded that "…bilateral HA amplification has advantages compared with unilateral amplification." In a related experiment, Van Wanrooij and Van Opstal [7], studied the acute effects of a monaural plug on directional hearing in the horizontal (azimuth) and vertical (elevation) planes of human listeners. They tested their subjects' localization behavior with rapid head orienting responses toward brief high-pass filtered (3 kHz; HP) and broadband (0.5–20 kHz; BB) noises, observing that the occlusion of one of the subject's ears with a plug "immediately degraded azimuth performance, as evidenced by a sound level– dependent shift ("bias") of responses

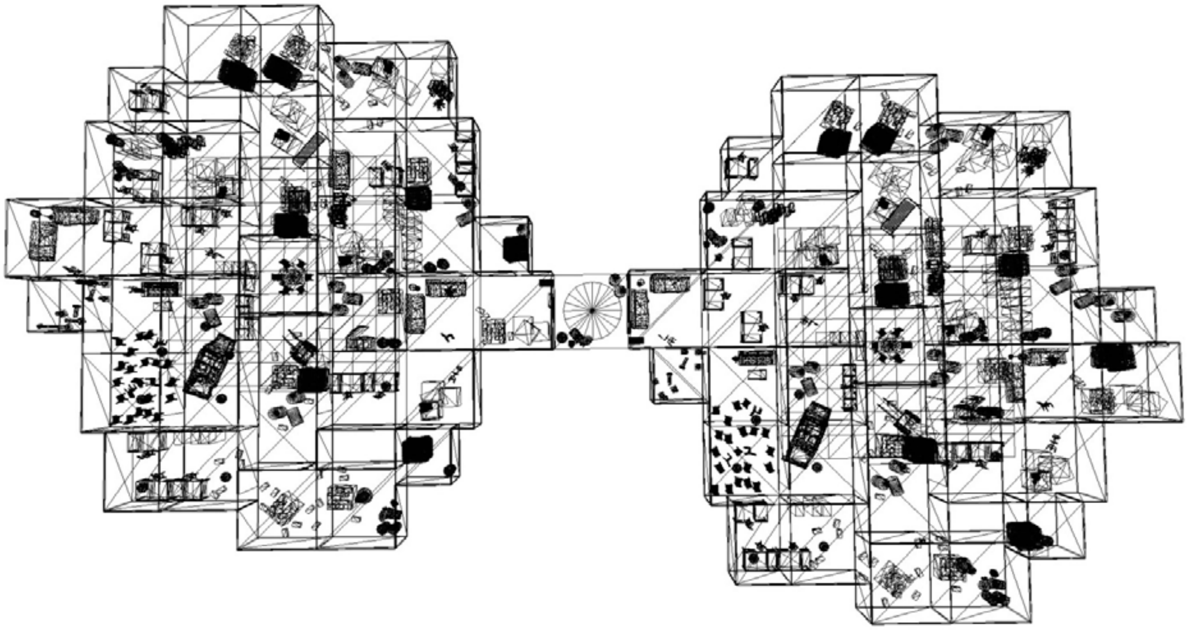

Figure 1: Symmetric mazes created in Unity® to test the two different types of auditory guidance. In each case, the navigator would start from the circle at the center of the figure.

contralateral to the plug," and that "also the elevation response components were affected by the plug." However, it has also been noticed that the incorporation of multi-sensory cues may help compensate for the degradation in localization ability due to the disruption of binaural hearing. Strelnikov, Rosito and Barone [8] tested the hypothesis that audiovisual integration can improve spatial hearing in monaural conditions when interaural difference cues are not available. They trained one group of subjects with an audiovisual task, where a flash was presented in parallel with a sound and another group in an auditory task, where only sound from different spatial locations was presented. From the results of their experiment, they concluded that "…cross-modal facilitation is highly important to improve spatial hearing in monaural conditions…"

This paper reports on the execution, results and analysis of an experiment in which we compared 2 versions of a VR environment which were different only in the type of audio simulation implemented in them.

II. Materials and methods

The objective of our experiment was to assess the difference in navigation proficiency displayed by human subjects in finding objects that acted as auditory beacons within a maze under 2 conditions. These conditions were: (1) when full binaural 3D sound (requiring head orientation tracking) was delivered, and (2) using an auditory signal that would be inversely proportional to the distance to the beacon, but would have the same amplitude in both ears, irrespective of orientation of the subject in the VR environment.

A. Virtual Test Environments

To test both auditory conditions, we created 2 symmetrical mazes using Unity® for which the primary layout differences were their entry points (e.g., one is entered from the East and the other one is entered from the West). The other significant difference between the mazes was in the type of sound guidance provided within each of them: one provided the sound of an acoustic beacon placed inside the maze as fully binaural 3D sound (hence this maze and the experimental condition that it implemented will be referred as '3D SOUND"). The other maze delivered the sound of the acoustic beacon within it with equal amplitude to both ears, which was inversely proportional to the distance to the beacon in the environment. This second maze and the experimental condition that it implemented will be referred as "2D SOUND" (since the amplitude does not depend on the navigator's head orientation, but only on the x and y coordinates of the navigator's position with respect to the beacon's position). Figure 1 shows a top view of both mazes. Figures 2 and 3 show examples of the rooms and the beacons in the mazes.

Figure 2: Example of one of the rooms in the virtual mazes

Figure 3: Example of one of the objects (guitar) that were used as acoustic beacons. Notice the soft green light from the beacon.

B. Virtual Reality Implementation Setup

Aiming for an immersive VR navigation experience, our subjects were fitted with an Oculus Rift (Development Kit version 2) VR headset and with Sennheiser HD280 Stereo headphones and a wireless keyboard.

The Oculus Rift provides stereoscopic display with a wide field of view and low latency on a low-persistence display. Further, the Oculus Rift contains an inertial measurement unit, involving accelerometers and gyroscopes, which provides continuous, real-time head orientation monitoring and drives the changes in the orientation view of the scene displayed, just as it happens for a human viewer in real life.

The subjects only needed to use two of the keys in the wireless keyboard for navigation: The "UP" arrow ([↑]) to advance, and the "DOWN" arrow ([↓]), to retrocede. Both keys were covered with patches of felt (of the type used as counterpart to Velcro patches), for easy tactile identification. Movement forward or backward would take place on a straight line defined by the point of view of the navigator into the virtual world. As such, to execute a turn in the virtual navigation a user would need to turn his / her body until the right direction is visualized "in front of him / her" and then advance by pressing the "UP" key. To facilitate these turning maneuvers, the subjects were seated on a stool capable of multiple 360 degree rotations and the cables connecting to the Oculus Rift and the headphones were dropped from above the subject using a microphone boom.

The stereo sound delivered through the headphones could be a fully binaural 3D sound (3D SOUND), or it could be set to be a sound with intensity that increases as the distance to the source decreases, but has the same level on both ears (2D SOUND). The difference can be set up in Unity® according to the "spread" property of the sound source used to implement the acoustic beacon: For 3D SOUND spread = 0, and for 2D SOUND spread = 180. Figure 4 shows the setup while a subject is navigating the virtual environment.

Figure 4: A subject performing virtual navigation. The visualization is performed with the Oculus Rift. The sound is delivered through stereo headphones and the subject advances or retrocedes by pressing arrow keys on the keyboard.

C. Experimental Subjects, Protocol and Task

Eight adult, healthy volunteers who reported normal hearing in both ears were involved in our experiment. Each subject would first be given an explanation of the virtual environment implementation that we used and was made aware of the fact that there would be two parts to the experiment, which would differ in the type of auditory guidance they would receive. Then subjects were given 5 minutes to navigate freely through each one of the 3D maze and the 2D maze. Then subjects would be given instructions to listen for the sound of an acoustic beacon (e.g., a telephone ringing, when the acoustic beacon was a telephone booth) and to navigate the maze to find the acoustic beacon. To differentiate the beacon from other objects in the virtual environment, it was also given the property to irradiate a soft green light. This light, however, would only be visible once the navigator reached into the room where the beacon was located. The subjects were told to navigate as to “crash” against the beacon to “acquire” it. Further, the subjects were told that once they “acquired” the first beacon, the second beacon (different object and correspondingly different sound) would appear somewhere else in the maze, and that they needed to find the second beacon. Similarly, after the second beacon was acquired, a third beacon would appear and would need to be acquired by the navigator. After the third beacon was acquired the experimental run would end. Each subject would complete a 3-beacon search in the 2D SOUND maze and another one in the 3D SOUND maze. The order of the mazes was varied at random for the different subjects, to minimize any bias caused by learning effects [9]. The objects used as acoustic beacons and their associated sounds were selected from the following: Piano (piano music); Telephone Booth (telephone ringing); Guitar (guitar music); Printer (dot-matrix printer sound) and Record Player (record music).

D. Statistical Design of Experiment

For this investigation, our hypothesis was:

> *Providing the virtual navigator with full binaural 3D sound will result in a significant improvement in the navigation performance within the virtual environment.*

Accordingly, the input variable to the experiment would by the SOUND TYPE (ST), with two levels: Fully binaural 3D guiding sounds, (3D) or guiding sounds that would have intensity inversely proportional to the distance from the navigator to the acoustic beacon, but would be delivered with the same intensity to both ears (2D).

To assess performance through quantitative, objective measures, these were the output variables:

T01 (Time for 1st search): Time measured from the start of the search run to the acquisition of the first acoustic beacon.

T12 (Time for 2nd search): Time measured from the acquisition of the first beacon to the acquisition of the second beacon.

T23 (Time for 3rd search): Time measured from the acquisition of the second beacon to the acquisition of the third and final beacon.

Therefore, the total time for each 3-beacon search would be (Ttot = Time for the complete search task):

$$Ttot = T01 + T12 + T23 \quad [1]$$

All of these times were logged to a text file, automatically, from the Unity® simulation.

In order to find out if our hypothesis could be supported we would analyze differences in Ttot, T01, T12 and T23 according to the 3D and 2D levels of the independent variable.

III. RESULTS

All search times (T01, T12, T23 and Ttot) were collected for all 8 subjects, performing the search task under both types of acoustic guidance (2D sound and 3D sound). Table I presents the averages for all four measurements under both conditions.

Table I. Average Search Times (seconds)

	2D-Sound		3D-Sound	
	Mean	**SD**	**Mean**	**SD**
T01	156.93	118.63	106.92	124.40
T12	63.08	29.46	40.22	15.45
T23	230.09	147.25	58.78	49.50
Ttot	450.10	257.72	205.92	114.75

The mean times required to complete the 3-beacon search is clearly longer for the 2D sound condition (450.10 seconds) than for the 3D sound condition (205.92 seconds). This same difference is appreciated in the mean times for each of the three segments of the search (T01, T12 and T23). However, it is necessary to conduct an appropriate statistical analysis to propose the verification of our initial hypothesis.

In this case we studied the significance of the difference of means for each of the time output variables (T01, T12, T23 and Ttot), with respect to the type of acoustic guidance (2D vs. 3D) provided. To that end we performed paired sample t-tests for each of the output variables. That is, we investigated the population of time differences (time under 2D condition – time under 3D condition) and tested the significance of the departure of the mean of differences from zero. The statistical analysis was carried out using SPSS [10].

In order to be able to apply the t-test properly, we verified that (for all T01, T12, T23 and Ttot) the populations of time differences were normal, according to the Kolmogorov-Smirnov and Shapiro-Wilk tests [11].

Table II displays the Paired Samples Statistics of the SPSS setup for our data. Table III shows the detailed results of the Paired Samples Test, from SPSS.

According to the results shown in Table III:

On average, participants performed the search for the first beacon with 3D sound (M=106.93, SE=43.98) faster than with 2D sound (M=156.93, SE=41.94). However, this difference was not significant $t(7)=1.406$, $p = 0.203$.

On average, participants performed the search for the second beacon with 3D sound (M=40.22, SE=5.46) significantly faster than with 2D sound (M=63.08, SE=10.42), $t(7)=2.528$, $p = 0.039$.

On average, participants performed the search for the third beacon with 3D sound (M=58.78, SE=17.50) significantly faster than with 2D sound (M=230.09, SE=52.06), $t(7)=3.277$, $p = 0.014$.

On average, participants performed the complete search task with 3D sound (M=205.92, SE=40.57) significantly faster than with 2D sound (M=405.10, SE=91.12), $t(7)=4.085$, $p = 0.005$.

IV. Discussion

The results of most of the statistical analyses, as well as the direct inspection of the data presented in Table I lead us to support our initial hypothesis. That is, providing the users with fully binaural 3D sound, in which the differences between the sounds delivered to the right and left ears will correspond to the orientation of the sound source with respect to the navigator's heading, improves the navigational performance.

There was, however, one of the output variables, T01, which is the time spent in the search of the first acoustic beacon, for which the difference was not found to be statistically significant. Furthermore, we can notice that the percentage of "excess" time required to complete each stage of the search under the 2D sound condition seems to have increased as the experiment went on, for each subject. That is, the "excess" time for T01 is (156.93 – 106.92)/106.92 = 0.4677, i.e., 46.77 %. In contrast, this "excess" time for T12 is (63.08 – 40.22)/40.22 = 0.5683, i.e., 56.83 % and, finally, for T23 the "excess" time is (230.09 – 58.78)/58.78 = 2.9144, i.e., 291.44 % .

We can appreciate that the difference in performance between the two sound delivery conditions became more accentuated as the experiment progressed. We believe that this might have been associated with the increasing frustration that some of the participants expressed while performing the searches with the 2D sound guidance.

Trying to understand the reasons for the improved navigational performance achieved under the 3D sound condition (and the subjective preference of the users for this method) we have distinguished a few possible factors.

Firstly, the 3D sound delivery provides important and useful information to the navigator at each point in which a path forward must be selected (e.g., when the user enters a room of the maze that has two exits), because it provides a sense of the approximate overall direction in which the navigation must proceed to reach the sound source. In contrast, the 2D sound delivery would only inform the navigator about how close to the sound source he/she currently is. Therefore, in order to "test" a potential way forward the navigator must actually move in a given direction and assess if this displacement might have brought him/her closer to the sound source.

Table II. Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T01_2D	156.9263	8	118.63022	41.94211
	T01_3D	106.9250	8	124.40107	43.98242
Pair 2	T12_2D	63.0788	8	29.46160	10.41625
	T12_3D	40.2163	8	15.44758	5.46154
Pair 3	T23_2D	230.0888	8	147.25458	52.06236
	T23_3D	58.7763	8	49.49888	17.50050
Pair 4	Ttot_2D	450.0963	8	257.72041	91.11792
	Ttot_3D	205.9175	8	114.75002	40.57026

In addition, the 3D sound condition provides the navigator with a means to "scan" different orientations, by simply turning his/her head while standing stationary, in order to enhance his/her judgment as to the precise direction of arrival of the spatialized sound. During the experiment we observed several of the subjects take advantage of this possibility when performing the search under the 3D sound condition.

Table III. Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	T01_2D - T01_3D	50.00125	100.61366	35.57230	-34.11387	134.11637	1.406	7	.203
Pair 2	T12_2D - T12_3D	22.86250	25.58411	9.04535	1.47365	44.25135	2.528	7	.039
Pair 3	T23_2D - T23_3D	171.31250	147.86491	52.27814	47.69434	294.93066	3.277	7	.014
Pair 4	Ttot_2D - Ttot_3D	244.17875	169.05893	59.77136	102.84195	385.51555	4.085	7	.005

V. CONCLUSIONS

Our experiment involved 8 human subjects performing virtual navigation under 2 different types of auditory guidance: 3D (full binaural spatialization) and 2D. In each case the subjects had to search for 3 auditory beacons, in a sequential fashion. We measured the amounts of time taken by each subject in reaching each of the 3 beacons, under the two types of auditory guidance. We observed that searches under the 2D condition took longer than those under the 3D sound condition. The difference, in percentage, between average times was smallest for the first search and largest for the third search. Our statistical analyses revealed that all the time differences between conditions (2D and 3D) were statistically significant, with the only exception of the difference for the search of the initial beacon.

These results provide strong confirmation of the importance that head tracking and the corresponding binaural implementation of spatialized 3D sound have for the realism and efficiency of contemporary immersive virtual environments.

In particular, we believe that a critical aspect of the benefits experienced by our subjects under the 3D sound condition was that they received immediate visual and auditory feedback whenever they simply turned their heads while remaining stationary. The Oculus Rift display would be continuously updated to reflect the change in the view achieved by the rotation and, simultaneously, the sound from the auditory beacon would be delivered with different intensities to the left and right ears. This scanning maneuver is one that we are accustomed to performing in the real world and can be highly beneficial when trying to localize a sound source. In contrast, under the 2D sound condition, only the visual output to the user would change if a subject turned his / her head while stationary. In addition to the unnatural feeling that this would evoke in the subject, this limitation yields a less effective navigation because it would force the subject to move forward, in a tentative direction, to update his / her estimation of the direction that may lead to the sound source.

Fortunately, new VR devices, such as the Oculus Rift, contain intrinsic instrumentation to provide continuous monitoring of the orientation of the user's head, which facilitates the delivery of virtual sounds on a fully-spatialized fashion.

ACKNOWLEDGMENTS

This work was sponsored by National Science Foundation grants HRD-0833093 and CNS-0959985.

REFERENCES

[1] Dinh, H. Q. Walker, L., Hodges, L., Song, C. and Kobayahi, A., "Evaluating the Inmportance of Multi-sensory Input on Memory and the Sense of Presence in Virtual Environmente." Proceedings of IEEE Virtual Reality, March 13-17, Houston, TX, pp. 222-228.

[2] Gröhn, M. , Lokki, T., Takala, T., "Comparison of auditory, visual, and audiovisual navigation in a 3D space," ACM Transactions on Applied Perception (TAP), v.2 n.4, p.564-570, October 2005.

[3] Loomis, J. M., Golledge, R. G., Klatzky, R. L. (1998)., "Navigation system for the blind: Auditory display modes and guidance." Presence: Teleoperators & Virtual Environments, 7, 193-203.

[4] Lokki, T., Grohn, M., Savioja, L., Takala, T., "A case study of auditory navigation in virtual acoustic environments," in Proc. ICAD 2000, Atlanta GA, Apr 2000, pp. 145–150.

[5] Zhou, Z., Cheok, A. D., Yang, X., Qiu Y., "An experimental study on the role of 3D sound in augmented reality environment." Elsevier, Interacting with Computers 16, (2004), pp. 1043-1068.

[6] Kobler S, Rosenhall U. Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. Int J Audiol. 2002;41:395-400.

[7] Van Wanrooij, M. M., & Van Opstal, A. J. (2007). Sound localization under perturbed binaural hearing. Journal of neurophysiology, 97(1), 715-726.

[8] Strelnikov, K., Rosito, M., & Barone, P. (2011). Effect of audiovisual training on monaural spatial hearing in horizontal plane. PloS one, 6(3), e18344.

[9] Lazar, J., Feng, J., and Hochheiser, H. Research methods in human-computer interaction. John Wiley & Sons Inc, 2009.

[10] Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). "SPSS explained". East Sussex, UK: Routledge.

[11] Field, Andy (2009), "Discovering statistics using SPSS", (3rd ed.) Sage, London, England (2009)