EFFECTS OF VISUAL AUGMENTATION ON THE MEMORY OF SPATIAL SOUNDS

Kyla A. McMullen

Clemson University
100 McAdams Hall
Clemson, SC 29634
kyla.mcmullen@gmail.com

Gregory H. Wakefield

University of Michigan
2260 Hayward St.
Ann Arbor, MI 48109
ghw@umich.edu

ABSTRACT

Spatial audio displays are created by processing digital sounds such that they convey a spatial location to the listener. These displays are used as a supplementary channel when the visual channel is overloaded or when visual cues are absent. This technology can be used to aid decision-makers in complex, dynamic tasks such as urban combat simulation, flight simulations, mission rehearsals, air traffic control, military command and control, and emergency services. Accurate spatial sound rendering is a primary focus in this research area, with spatial sound memory receiving less attention. The present study assesses the effects of visual augmentation on spatial sound location and identity memory. The chosen visual augmentations were a Cartesian and polar grid. The work presented in this paper discovered that the addition of visual augmentation improved location and identity memory without degrading search time performance.

1. INTRODUCTION

Digital sounds can be processed such that auditory cues are created that convey spatial location within a virtual auditory environment (VAE). In this environment, locations are indicated by the perceived position of the processed sound sources. These spatial audio displays are advantageous in that they enable "eyes-free" interaction, which can allow a system designer to utilize multiple sources of information without needing to pay visual attention to a screen [1]. The ability to perceive virtually placed sounds while receiving real-time spatialized audio cues, is a relatively new concept that may not be natural to most listeners.

A spatial audio system uses a collection of customized digital filters for each listener and outputs the spatial sound over headphones. These filters, known as Head-Related Transfer Functions (HRTFs), differ across individuals as a function of head shape, placement of the pinna, and the shape of the pinna and ear canal. Work by Roginska et al. [2] and McMullen et al. [3] demonstrate that listeners can select from a database of HRTFs ones that function similarly to their own. This selection process removs the need for expensive individual measurement. Furthermore, Roginska et al. [4] demonstrates that listeners can use HRTFs to navigate and locate sounds in a VAE.

The relative ease by which people navigate through space on a daily basis reflects the tight coupling of the human auditory and visual systems in their support of spatial perception. Due to the complementary nature of each sensory modality, when one or the other is compromised, individuals can accommodate, to a limited extent, by learning to use the uncompromised system in new ways. This is particularly evident for the blind that benefit considerably in both local and global navigation through extensive training.

With recent advances in the quality and portability of video and audio systems, the emerging challenge of pervasive virtual reality (VR) is to build systems that leverage the tight perceptual coupling of vision and audition while meeting the specific demands of the task at hand. While much is already known about how well users navigate VR worlds using visual information alone, we are only beginning to understand navigation using the auditory channel due to the lag in the development of wearable audio rendering systems. Indeed, in some applications where visual cues are degraded or the direct sensation of the world may not be possible or desirable, optimally supporting a "blind" VR experience through perceptually well-matched audio is key [5].

2. RELATED WORK

The auditory channel can be used as a supplementary channel when the visual channel is overloaded or when visual cues are absent. It is generally agreed that the addition of sound to visual information aids data representation by adding an additional dimension and providing representation redundancy to confirm visual details [6, 7, 8, 9, 10].

Spatial audio displays enhance visual search performance by creating a multimodal system. Several researchers have demonstrated the effectiveness of spatial auditory cues in enhancing visual search performance [11, 12, 13, 14]. The auditory channel can monitor objects that may not have been perceived by the limited visual channel. This proposed multimodal system would have many advantages such as: error prevention, interface robustness, error correction/recovery, increasing communication bandwidth, and providing alternate communication methods [15].

This technology can be used to aid decision-makers in complex, dynamic tasks such as urban combat simulation, flight simulations, mission rehearsals, air traffic control, military command and control, and emergency services. For example, Bastide [16] uses spatial sound to create a multimodal command and control interface for the Rafale aircraft. Additionally, Nguyen uses spatial audio to provide assistance when escaping to an emergency exit [17]. Due to the fact that these kinds of systems that use spatial audio to convey location, it is important to study how people recall the positions of spatially rendered sounds.

Although spatial audio displays are not present in current military systems, several applications have been suggested. These include including monitoring multiple radio communications channels, navigating waypoints, locating systems, warning of threats and malfunctions, and teleoperation of unmanned vehicles [18]. In cockpit applications with helmet- or head-mounted visual displays with a limited field of view, spatial audio can be used to direct the attention of the pilot to critical events occurring outside the visual FOV. Haas et al. [19] investigated the use of spatial auditory displays in helicopter cockpit radio communications tasks. The United States Air Force also experimented with the use of spatial auditory displays to provide fixed wing aircraft with waypoint information [20].

The potential uses of spatial audio technology are limitless, however the field would greatly benefit from research investigating the memory of localized spatial sounds. Memory performance in spatial audio systems has received limited attention even though it is a critical factor in predicting and determining action, in critical tasks.

It is possible that a visual augmentation, such as a coordinate system, would help listeners remember the positions of auditory objects. For example, in the visual domain, Leifert [21] discovered that the addition of grid lines provides a structural anchor that helps participants to remember spatial locations, no matter how the objects are aligned. Perhaps the same holds true in the auditory domain. Leifert's work also discovered that the grid lines has a contrary effect on content (or identity) memory.

If the addition of a visual coordinate system improves performance, system designers would be encouraged to consider incorporating this visual augmentation into the interface. On the other hand, the additional cognitive load of interacting with a visual reference frame while monitoring an auditory object may degrade perception and memory performance. The present work examines how auditory and visual cues interact during the recall of auditory spatial objects. The present study also seeks to determine the effects of using an advanced visual cue during exploration and assesses its effects on recall.

To test the effects of visual augmentation, listeners explored a five-source auditory environment that was augmented with Cartesian or polar reference frame. Performance is assessed in terms of location and content memory, as spatial information is not processed independently of other sensory perception. In addition, exploration time is also measured. Both metrics are compared to the participant's own performance, in a condition without visual augmentation.

3. METHOD

3.1. Participants

Five listeners (2 women and 3 men) who were undergraduate students participated in the current experiment. Each participant was screened to have normal hearing, through audiogram measurement. Each listener was experienced in a spatial-audio listening task, having participated in experiments in a prior HRTF-customization [3] and sound source localization training experiments. The experiment required about 2.2 hours of listening and was completed in one session. Participants were paid $10 per hour. Before taking part in the study, each participant gave his or her consent by reading and signing a consent form.

3.2. Apparatus

Experiments were conducted in a Tracoustics soundproof booth in the Computer Science and Engineering Building at the researchers' university. Each participant was seated at a table in front of an iMac desktop computer. An Apogee Duet audio interface was used to generate the left and right channels of the audio signal, which was delivered to the listener over Beyerdynamic headphones. As in [4], a real-time spatial auditory system programmed in MATLAB was used. The system spatialized the auditory stimuli using each participant's customized HRTF, which was created using the procedure described in [3]. The controller sampled the participant's position and orientation at a rate of 10 Hz. To facilitate a real-time double buffering audio scheme, a timer was used to generate a new frame of audio corresponding to the participant's position within the VAE. The timer called routines to read the sound file from disk and convolve the audio input using the orientation-adjusted interpolated HRTFs. Audio was controlled using the PsychToolbox [22] extension of OpenAL [23] by double buffering. The timer also queried the OpenAL sound source to determine if one of the buffers had finished playing, so that the next frame of audio could be loaded into the buffer, to be played after the current buffer. Interpolation of the HRTFs was implemented by constructing the minimum-phase impulse response of a system whose magnitude spectrum is determined from a log mixture of the adjacent measured HRTFs (sampled every 10 degrees) and convolving the result with an all-phase system using a fractional-delay method. An inverse-squared law was used to attenuate the sounds as the listener moved through the VAE. Standard mouse to screen cursor mapping was used.

The Cartesian, polar, and standard interfaces are displayed in Figure 1. Navigation was facilitated from a bird's-eye-view perspective. The listener's position in the environment is indicated by the location of the red circle, which serves as the avatar. Their orientation is indicated by the position of the red line (or nose). Clicking or dragging the mouse to a new location moves the position of the onscreen avatar. The keyboard's right and left arrows were used to control the yaw of the avatar's head by rotating its orientation in steps of two degrees.

The interface was presented from a bird's-eye-view perspective, as motivated by the results of [24] which found that there was no significant difference in performance when listeners located sounds on the horizontal plane (first person perspective) as compared to localization on the vertical plane (bird's-eye-view perspective).

3.3. Stimuli

Table 1: Environmental sounds and their labels

Sound	Labels
Drumsticks striking a drum at regular intervals	Drums
Computer generated electronic noises	Electronic
A river flowing rapidly	River
Crickets chirping	Crickets
Typewriter keys being pressed	Typewriter

The experimental stimuli consisted of five distinct environ-

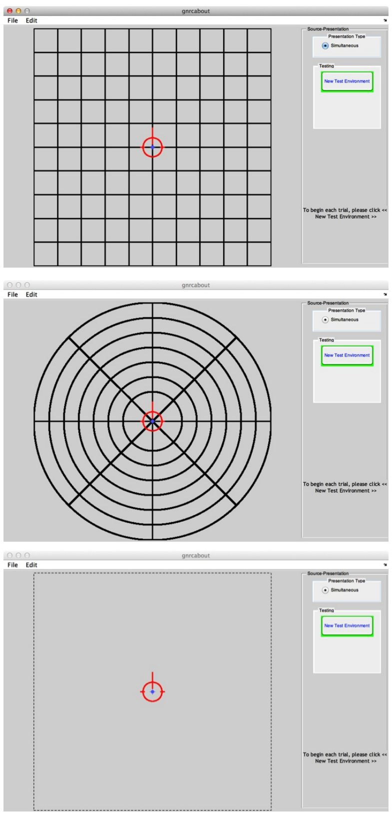

Figure 1: Visual augmentations used in the present study. Cartesian interface (top), Polar interface (middle), No visual augmentation (bottom)

mental sounds (see Table 1). The sounds were chosen from the BBC Sound Effects Library [25] based upon their ability to be perceptually segregated as five individual sounds and because of their distinct spectro-temporal patterns. Each signal also contained broadband energy distribution and transients, to support binaural localization [26, 27]. Sounds were intuitively chosen to be easily distinguished from one another. Each sound was between 23 and 80 seconds long, and repeated continuously. Stimuli were presented at an audible level, as adjusted by the participant.

3.4. Procedure

Before beginning each condition, participants completed a two-phase training procedure to orient themselves to the auditory environment as realized through the particular visual augmentation. In the training, the listener completed a minimum of 20 axial training tasks in which they had to accurately localize a sound that was either directly in front, back, to the left, or to the right of them. Training was performed in this manner as a result of the findings of Roginska et al. [4] that identified front-back confusion as a major challenge in virtual sound localization. Roginska et al. also found that when localizing sounds, listeners often try to achieve equal sound volumes in both ears, thus necessitating the need to practice balancing sound levels in both ears through left/right practice. Training continued until the participant reached a predetermined accuracy criteria for five consecutive trials. Random-placement training followed axial training, and in a similar fashion, training continued until the participant reached a predetermined accuracy criteria for five consecutive trials. After both of the training phases were completed, the participant began the experiment.

As illustrated in Figure 2, listeners began each experimental trial facing forward in the middle of the silent VAE. After pressing a button, the five stimuli played simultaneously at randomly chosen locations within a unit circle around the listener. The locations of each sound were chosen pseudo-randomly on each trial so that each sound was spaced at least 30° apart and separated by a distance that equaled one-third of the diameter of the auditory space.

On each trial, the participant explored the auditory environment by moving the cursor on the computer screen using a free-search procedure with the goal of finding and memorizing the locations of the five sound sources. When the participant acquired the spatial configuration of the sources, they pressed the space bar to stop playing the sources.

Next, the listener marked the locations where they remembered hearing the sound sources by clicking the mouse, which placed a blue dot at the mouse's current position on the screen. After all of the locations had been marked, the participant labeled the identities of the sound sources they had previously marked using a drop down list.

3.5. Design

Listeners completed twenty trials of two experimental conditions, half beginning with the polar interface and the other half beginning with the Cartesian grid interface, in a balanced design. Each participant's training entailed completing a self-guided training procedure before beginning each condition. The performance data for the collected for each participant in the experimental manipulations was compared to a previous task that was administered in the same manner for twenty trials without visual augmentation.

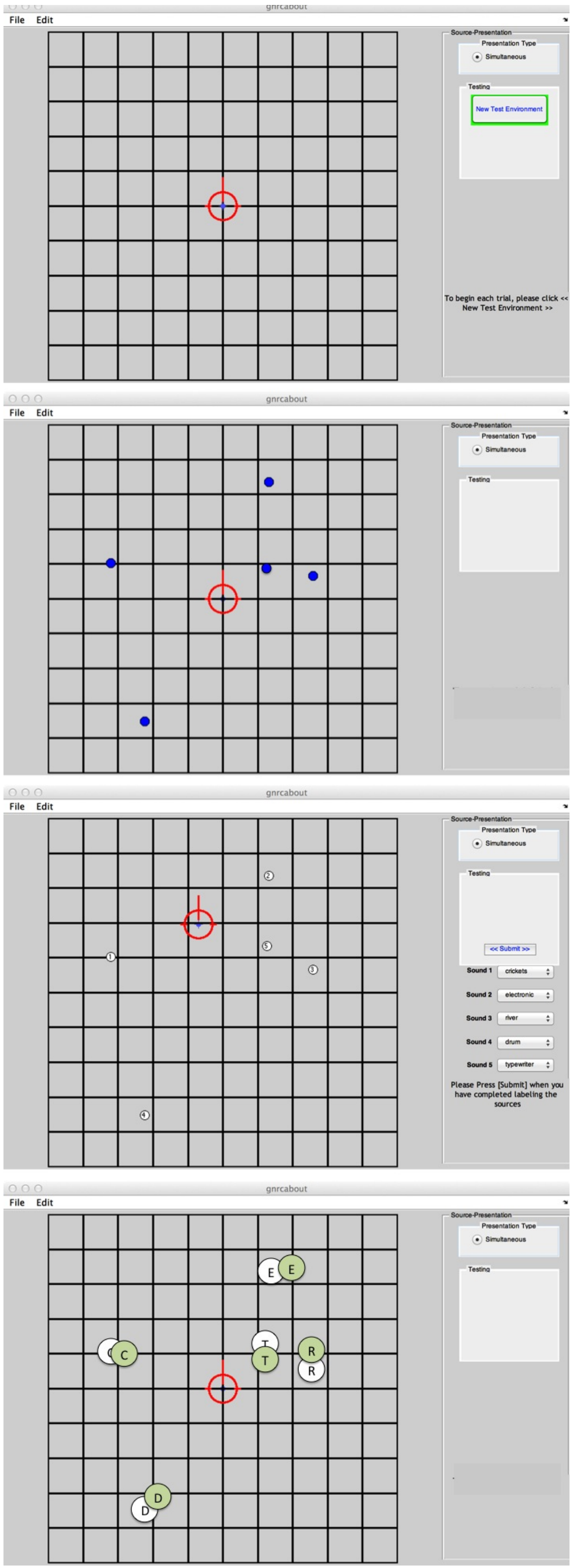

Figure 2: User Interaction with a Cartesian visual augmentation. The red circle represents the listeners position and the long red line intersecting the circle represents the heading from a top-down perspective. The user begins by exploring the environment while memorizing the positions of each sound source (top panel). Next, the sounds are removed and the user marks the locations of the recalled sounds (second panel). The locations change to numbers and the user selects the identity of each numbered sound (third panel). Finally, feedback is given(last panel - feedback enlarged for publication readability)

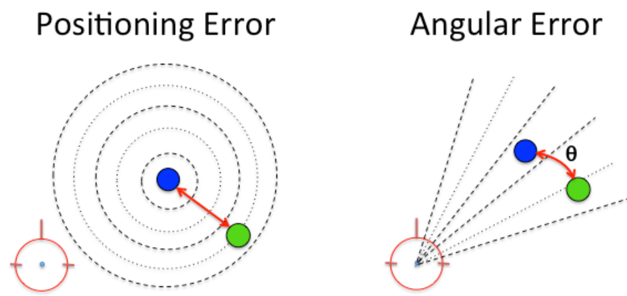

Figure 3: Positioning error (left) and angular error (right). In both figures, the blue circle represents the true location of the sound source and the green circle represents the location of the source as marked by the user.

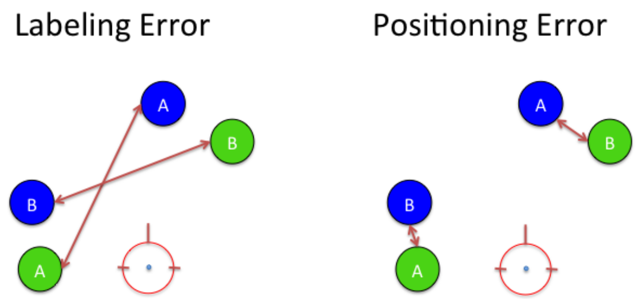

Figure 4: The distinction between labeling error (left) and Positioning error (right). In both figures, the blue circle represents the true location of the sound source and the green circle represents the location of the source as marked by the user. This metric assesses the recall accuracy of the correct sound identity at each marked location.

As is standard in psychoacoustic experiments, a large amount of data was collected from a small and well trained selection of listeners. The accuracy results of the present experiment were evaluated by using three performance measures: *positioning error*, *angular error*, and *labeling error* that are illustrated in Figures 3 and 4.

Positioning error was defined as the straight-line distance between the true and marked sound location. Positioning error was calculated in term of R_{Max}, which represents the maximum distance from which a sound source could be located with respect to the listener's position in the center of the interface. R_{Max} can range from 0 to 1 inclusively. *Angular error* was defined as the unsigned angular difference between the actual and perceived sound source. The sound's angle was defined as the angle between the interaural axis and a line originating at the center of the interface that intersects the sound's location. *Labeling error* describes the difference between a sound's actual location and its recalled location. This measure was introduced to distinguish the difference (if any) between recalling the configuration of the environment (positioning error) as compared to correctly recalling each source, identified in its correct position. Multidimensional scaling was used to determine a linear transformation of the user's marked locations that best conformed to the real point configuration. Dissimilarity between the transformed configuration and the actual locations of the sound sources was measured. This was done to account for any

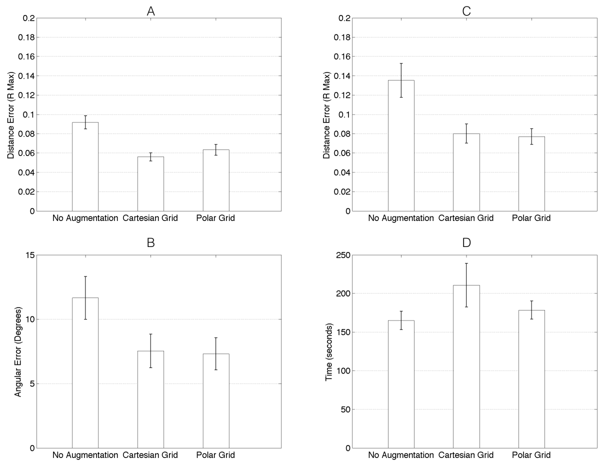

Figure 5: Effects of visual augmentation on positioning(A), angular(B), labeling(C), and time (D) accuracy. Along the abscissa are the augmentation conditions and along the ordinate is the positioning error. The error bars show the 95% confidence interval.

overall configuration scaling or shifting while marking the source positions. Exploration time was also measured by assessing the difference between start and stop time for each listener to navigate the environment while learning the source positions. The timer began when the listener entered the environment by pressing "New Test Environment"and ended after they pressed the spacebar, indicating that all sounds and locations had been memorized.

Each analysis displays the results, collapsed across subjects. Statistical significance was determined at the $p < 0.05$ level.

4. RESULTS

4.1. Accuracy

Figure 5 (A) shows the mean positioning error during the three visual augmentation conditions. A significant difference in positioning error was observed as an effect of visual augmentation [$F_{2,1497}$=43.57, $p<0.05$]. A Tukey Least Significant Difference (LSD) multiple comparison test showed that positioning error was significantly lower with either visual augmentation.

Similar results were observed in the comparison of angular error. Figure 5 (B) shows the mean angular error during the three visual augmentation conditions. A significant difference in angular error was observed as an effect of visual augmentation [$F_{2,1497}$=2992.3, $p<0.05$]. A Tukey LSD multiple comparison test showed that angular error was significantly lower when either visual augmentation was used, as compared to no visual augmentation.

As in the previous two analyses, the labeling accuracy follows a similar trend. Figure 5 (C) shows that there was a significant difference in labeling error as an effect of visual augmentation [$F_{2,1497}$=26.22, $p<0.05$]. A Tukey LSD multiple comparison test showed that labeling error was significantly lower when either visual augmentation was used, as compared to no visual augmentation.

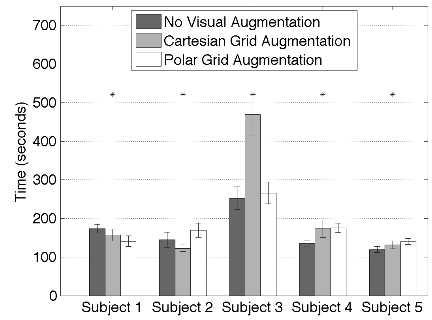

Figure 6: Exploration time as affected by visual augmentation. Along the abscissa are the subjects and along the ordinate is the total exploration time. The error bars show the 95% confidence interval.

4.2. Exploration Time

Figure 5 (D) shows the average time listeners explored the environment using a polar, Cartesian, or no reference frame. As in the previous analyses, there was a significant difference in exploration time as an effect of visual augmentation [$F_{2,297}$=6.02, $p<0.05$]. A Tukey LSD multiple comparison test showed that exploration time was higher in the Cartesian grid condition than the other two conditions.

The data was further broken down by subject (Figure 6) and a 2-way ANOVA was performed. It can be observed that subject 3 needed significantly more time to explore the Cartesian environment. The second analysis (omitting the outlying data from Subject 3) still revealed that visual augmentation significantly affected each subject's exploration time [$F_{2,228}$=4.35, $p<0.05$].

4.3. Visual Memory Comparison

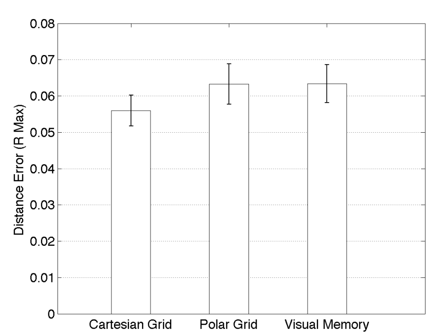

Figure 7: Comparison of auditory recall performance to visual recall performance. Along the abscissa are the conditions and along the ordinate is the total exploration time. The error bars show the 95% confidence interval.

Figure 7 shows the comparison study in which the same listeners were presented in the same manner with circles of various colors on a screen, and were asked to recall the spatial locations of the circles that had been presented. The results show that there was not a significant difference in positioning error, between the three treatments [$F_{2,1497}$=2.75, p=0.06].

5. DISCUSSION AND CONCLUSION

Participants exhibited significantly higher positioning, angular and labeling accuracy when an auditory environment was augmented with a Cartesian or polar reference frame. No effect was seen across the two types of visual augmentations. This result is interesting in light of Leifert's work [21] that found that grid lines have a contrary effect on content (or identity) in visual memory. This finding suggests that there may be a difference between recalling the identity of a visual object as compared to a sound.

The results suggest that the addition of visual augmentation can lower angular error. In virtual systems utilizing non-individualized HRTFs, localization error has been found to be slightly more than that which was observed in the present study, ranging between 10 degrees and 30 degrees, depending on sound direction [28].

It is interesting to note that some researchers have found that visual objects are more easily remembered with a salient visual reference frame. For example, Carlson[29] and Kelly et al. [30] found that incorporating this type of salient cue enables a participant to form a more accurate mental spatial representation of the environment. Tversky [31] states that in environments where spatial orientations are difficult to remember, heuristics, such as reference points, are used to anchor figures to locations, making them easier to remember. Von Wright et al. [32] found that the use of external spatial frames of reference aids in the encoding of attributes of visual objects in young children. On the other hand, Albert et al. [33] indicates that visual reference frames have little, if any, impact on the acquisition of spatial relationships among visually displayed objects.

However, in the auditory domain, Zahorik [34] notes that it is important to recognize the contributions of non-acoustical (visual) factors that affect the perception of auditory space. Although the role of visual reference frames in visual memory is debatable, the present work suggests that visual reference frames support the acquisition of auditory spatial locations and identities. Furthermore, the results indicate that listeners perform just as well in recalling the location of objects in visual search as compared to auditory search.

Participants exhibited significantly higher recall accuracy when any visual augmentation was used. There was no difference in accuracy between the two visual augmentations. One might have expected angular accuracy to be significantly higher in the polar augmentation condition, due to the structure of the reference frame. Perhaps reference frames are used as general-purpose landmarks. Their shape may not influence sound search and recall. In the future, this line of inquiry will examine zenith estimation in addition to azimuth and elevation to determine if the same effects are observed. Additionally, further investigation is being conducted to assess the effects of practice on user performance. It may be possible that users were more familiar with the task, as a result of having participated in a similar experiment prior to the current procedure.

Interestingly, one participant spent significantly more time exploring the Cartesian augmented environment, followed by the polar augmented environment. The higher exploration time may have been a result of participants counting the boxes in an attempt to memorize the exact coordinate or square in which each source was located.

Findings from this experiment suggest that VAE designers should incorporate visual reference frames into their visual positional display of auditory spaces. However, they must note that an increase in search time may be observed if the system uses an augmentation that encourages the operator to memorize features of the reference frame. If the operator is performing a task that is not time-sensitive, any frame could be used.

6. REFERENCES

[1] G. N. Marentakis and S. A. Brewster, "Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 359–368.

[2] A. Roginska, G. H. Wakefield, and T. S. Santoro, "User selected hrtfs: Reduced complexity and improved perception," Undersea Human Systems Integration Symposium, Tech. Rep., July 2010.

[3] K. McMullen, A. Roginska, and G. Wakefield, "Subjective selection of head-related transfer functions (hrtf) based on spectral coloration and interaural time differences (itd) cues," in *Audio Engineering Society Convention 133*, 2012.

[4] A. Roginska, G. H. Wakefield, T. S. Santoro, and K. A. McMullen, "Effects of interface type on navigation in a virtual spatial auditory environment," in *Proceedings of the International Conference on Auditory Displays*, 2010.

[5] E. M. Wenzel, F. L. Wightman, and S. H. Foster, "A virtual display system for conveying three-dimensional acoustic information," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 32, pp. 86–90(5), 1988. [Online]. Available: http://www.ingentaconnect.com/content/hfes/hfproc/1988/00000032/00000002/art00018

[6] D. B. Anderson and M. A. Casey, "The sound dimension," *Spectrum, IEEE*, vol. 34, no. 3, pp. 46–50, 1997.

[7] T. Hermann and H. Ritter, "Sound and meaning in auditory data display," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 730–741, 2004.

[8] H. G. Kaper, E. Wiebel, and S. Tipei, "Data sonification and sound visualization," *Computing in science & engineering*, vol. 1, no. 4, pp. 48–58, 1999.

[9] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda. international community for auditory display," *Retrieved July*, vol. 27, p. 2007, 1999.

[10] R. D. Shilling and B. Shinn-Cunningham, "Virtual auditory displays," DTIC Document, Tech. Rep., 2000.

[11] B. Elias, "The effects of spatial auditory preview on dynamic visual search performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 40, no. 23. SAGE Publications, 1996, pp. 1227–1231.

[12] G. E. Fujawa and T. Z. Strybel, "The effects of cue informativeness and signal amplitude on auditory spatial facilitation

of visual performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 41, no. 1. SAGE Publications, 1997, pp. 556–560.

[13] D. R. Perrott, J. Cisneros, R. L. McKinley, and W. R. D'Angelo, "Aurally aided detection and identification of visual targets," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 39, no. 1. SAGE Publications, 1995, pp. 104–108.

[14] T. Z. Strybel, J. M. Boucher, G. E. Fujawa, and C. S. Volp, "Auditory spatial cueing in visual search tasks: Effects of amplitude, contrast and duration," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 39, no. 1. SAGE Publications, 1995, pp. 109–113.

[15] P. R. Cohen and D. R. McGee, "Tangible multimodal interfaces for safety-critical applications," *Communications of the ACM*, vol. 47, no. 1, pp. 41–46, 2004. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Tangible+Multimodal+interfaces+for+Safety-Critical+Applications#0

[16] R. Bastide, D. Navarre, P. Palanque, A. Schyn, and P. Dragicevic, "A model-based approach for real-time embedded multimodal systems in military aircrafts," in *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 2004, pp. 243–250.

[17] V. Y. Nguyen, "Audible assistance on escaping to an emergency exit: A comparison of a manipulated 2d and a 3d audio model," 2004.

[18] J. Y. Chen, E. C. Haas, K. Pillalamarri, and C. N. Jacobson, "Human-robot interface: Issues in operator performance, interface design, and technologies," Army Research Laboratory, Tech. Rep., 2006.

[19] E. C. Haas, C. Gainer, D. Wightman, M. Couch, and R. Shilling, "Enhancing system safety with 3d audio displays," in *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, 1997.

[20] R. L. McKinley, M. A. Erickson, and W. R. D'Angelo, "3-dimensional auditory displays: Development, applications, and performance." *Aviation, Space, and Environmental Medicine*, 1994.

[21] S. Leifert, "The influence of grids on spatial and content memory," in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*. ACM, 2011, pp. 941–946.

[22] M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, and C. Broussard, "What's new in psychtoolbox-3," *Perception*, vol. 36, no. 14, pp. 1–1, 2007.

[23] G. Hiebert, "Openal 1.1 specification and reference," 2005.

[24] K. A. McMullen and G. H. Wakefield, "Effects of plane mapping on sound localization in a virtual auditory environment," in *HCI International 2013-Posters' Extended Abstracts*. Springer, 2013, pp. 357–361.

[25] BBC, *The BBC Sound Effects Library*. Princeton, NJ: Princeton, NJ: Films for the Humanities and Social Sciences, 1991, vol. 1-40.

[26] D. Begault, "3-d sound for virtual reality and multimedia," 2000.

[27] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT Press, 1983.

[28] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using non-individualized head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[29] L. Carlson, "On the "whats" and "hows" of "where": The role of salience in spatial descriptions," in *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008, vol. 5248, pp. 4–6.

[30] J. Kelly, M. Avraamides, and T. McNamara, "Reference frames influence spatial memory development within and across sensory modalities," in *Spatial Cognition VII*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, pp. 222–233.

[31] B. Tversky, "Cognitive maps, cognitive collages and spatial mental models," in *Spatial Information Theory: A Theoretical Basis for GIS – Proceedings of COSIT'93*, A. Frank and I. Campari, Eds. Berlin: Springer, 1993, pp. 14–24.

[32] J. M. Von Wright, P. Loikkanen, and P. Reijonen, "A note on the development of recall of spatial location." *British Journal of Psychology*, vol. 69, no. 2, pp. 213–216, 1978.

[33] W. S. Albert, R. A. Rensink, and J. M. Beusmans, "Learning relative directions between landmarks in a desktop virtual environment," *Spatial Cognition and Computation*, vol. 2, pp. 131–144, 1999.

[34] P. Zahorik, "Auditory display of sound source distance," *Source*, pp. 1–7, 2002. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.3154&rep=rep1&type=pdf